Technical Report 1283

Tier One Performance Screen Initial Operational Test and Evaluation: Early Results

Deirdre J. Knapp (Ed.)
Human Resources Research Organization

Tonia S. Heffner and Leonard White (Eds.)
U.S. Army Research Institute

April 2011

United States Army Research Institute
for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical review by

Sharon Ardison, U.S. Army Research Institute
J. Douglas Dressel, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
April 2011	Final	August 2009 to August 2010

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Tier One Performance Screen Initial Operational Test and Evaluation: Early Results	W91WAW-09-C-0098
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Deirdre J Knapp (Ed.)(Human Resources Research Organization), Tonia S. Heffner and Leonard White (Eds.)(U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 329
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Human Resources Research Organization 66 Canal Center Plaza, Suite 700 Alexandria, Virginia 22314	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-RS 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Technical Report 1283

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter Expert POC: Dr. Tonia Heffner

14. ABSTRACT *(Maximum 200 words)*:

Along with educational, medical, and moral screens, the U.S. Army uses a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT) to select new Soldiers. Although the AFQT is useful for selecting new Soldiers, other personal attributes are important to Soldier performance and retention. Based on the U.S. Army Research Institute's (ARI) investigations, the Army selected one promising measure, the Tailored Adaptive Personality Assessment System (TAPAS), for an initial operational test and evaluation (IOT&E), beginning administration to applicants in 2009. Criterion data are being collected at 6-month intervals from administrative records, from Initial Military Training (IMT), and from schools for eight military occupational specialties (MOS) and will be followed by two waves of data collection from Soldiers at first unit of assignment. This is the first of six planned evaluations of the IOT&E. This report documents the early analyses from a small sample of Soldiers who completed the TAPAS and completed IMT. Similar to prior experimental research, our early evaluation suggests that several TAPAS scales significantly predicted a number of criteria of interest, indicating that the measure holds promise for both selection and classification purposes.

15. SUBJECT TERMS

behavioral and social science, personnel, manpower, selection and classification

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	82	Ellen Kinzer Technical Publications Specialist (703) 545-4225

Standard Form 298

Technical Report 1283

Tier One Performance Screen Initial Operational Test and Evaluation: Early Results

Deirdre J. Knapp (Ed.)
Human Resources Research Organization

Tonia S. Heffner and Leonard White (Eds.)
U.S. Army Research Institute

Personnel Assessment Research Unit
Michael G. Rumsey, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

April 2011

Army Project Number
622785A790

Personnel, Performance
and Training Technology

Approved for public release; distribution is unlimited.

ACKNOWLEDGEMENTS

There are individuals not listed as authors who made significant contributions to the research described in this report. First and foremost are the Army cadre who support criterion data collection efforts at the schoolhouses. These noncommissioned officers (NCOs) ensure that trainees are scheduled to take the research measures and provide ratings of their Soldiers' performance in training. Thanks also go to Dr. Brian Tate and Ms. Sharon Meyers (ARI) and Mr. Doug Brown, Ms. Ashley Armstrong, Mr. Blane Lochridge, and Ms. Mary Adeniyi (HumRRO) and Mr. Jason Vetter (Drasgow Consulting Group) for their contributions to this research effort.

We also want to extend our appreciation to the Army Test Program Advisory Team (ATPAT), a group of senior NCOs who periodically meet with ARI researchers to help guide this work in a manner that ensures its relevance to the Army and help enable the Army support required to implement the research. Members of the ATPAT are listed below:

CSM JOHN R. CALPENA
CSM BRIAN A. HAMM
CSM JAMES SHULTZ
CSM (R) CLARENCE STANLEY
SGM (R) DANIEL E. DUPONT SR.
SGM JOHN EUBANK
SGM KENAN HARRINGTON
SGM THOMAS KLINGEL
SGM(R) CLIFFORD MCMILLAN
SGM HENRY C. MYRICK
SGM GREGORY A. RICHARDSON
SGM RICHARD ROSEN
SGM MARTEZ SIMS
SGM BERT VAUGHAN
1SG ROBERT FORTENBERRY
MSG JAMES KINSER
MSG DARRIET PATTERSON
MSG ROBERT D. WYATT
SFC QUINSHAUN R. HAWKINS
SFC WILLIAM HAYES
SFC STEVEN TOSLIN
SFC KENNETH WILLIAMS

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: EARLY RESULTS

EXECUTIVE SUMMARY

Research Requirement:

In addition to educational, physical, and moral screens, the U.S. Army relies on a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers into the Army. Although the AFQT has proven to be and will continue to serve as a useful metric for selecting new Soldiers, other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Campbell & Knapp, 2001; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009, 2010; Knapp & Tremble, 2007). Based on ARI's research, the Army selected one particularly promising measure, the Tailored Adaptive Personality Assessment Screen (TAPAS), as the basis for an initial operational test and evaluation (IOT&E) of the *Tier One Performance Screen*.[1] TAPAS capitalizes on the latest in testing technology to assess motivation through the measurement of personality characteristics.

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering the TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB) at Military Entrance Processing Stations (MEPS). The WPA will be introduced for applicant testing in CY2011. The plan is to continue administration as part of the IOT&E through FY 2013. Criterion data are being collected from administrative records at 6-month intervals. As part of the IOT&E, initial military training (IMT) criterion data are being collected at schools for eight military occupational specialties (MOS) and will be followed by two waves of data collection from Soldiers once they are in their units.

Procedure:

The typical delay between pre-enlistment testing and when individuals actually enter the Army resulted in small samples on which to conduct validation analyses. Specifically, whereas there were almost 54,000 applicants who took the TAPAS, of which just over 24,000 signed an enlistment contract, the August 2010 database only has administrative criterion data on roughly 3,500 Soldiers and IMT data on fewer than 400. Thus, the selection and classification-oriented analyses reported here must be viewed with considerable caution.

To compare the internal and external psychometric properties of TAPAS across versions (nonadaptive or "static", and adaptive) and settings (research vs. IOT&E), we conducted a series of analyses. In this IOT&E, three versions of TAPAS were administered: a 13-dimension, 104-item adaptive test, a 15-dimension, 120-item nonadaptive test, and a 15-dimension, 120-item adaptive test. An effort was made to enhance consistency across test versions by maintaining a

[1] The Work Preferences Assessment (WPA) was identified as another promising measure to be included in the IOT&E. The WPA asks respondents their preference for various work activities and environments.

common set of dimensions and using the same matching constraints for item construction. However, equivalence was not possible due to the differences in content, length, and item selection methods.

Our approach to analyzing the TAPAS' incremental predictive validity was consistent with previous evaluations of this measure and similar experimental non-cognitive predictors (Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009; 2010). In brief, this approach involved testing a series of hierarchical regression models, regressing each criterion measure onto Soldiers' AFQT scores in the first step, followed by their TAPAS scale scores in the second step. When the TAPAS scale scores were added to the baseline regression models, the resulting increment in the multiple correlation (ΔR) served as our index of incremental validity.

Given our very low MOS-specific sample sizes, we were unable to conduct planned analyses to examine classification efficiency at this time. Instead, we examined cross-MOS differences in TAPAS score profiles and predictive validity estimates to get an idea of TAPAS' potential as a classification tool. Specifically, we computed the overall average root mean squared difference (RMSD) in TAPAS scale scores across MOS. Similar to the selection analyses, cross-MOS differences in predictive validity estimates were measured by computing an average RMSD in these estimates among the MOS sampled.

Findings:

The results of the selection-oriented analyses suggest that the individual TAPAS scales significantly predict a number of criteria of interest. Most notably, the Physical Conditioning scale predicted Soldiers' self-reported Army Physical Fitness Test (APFT) scores, number of restarts in training, adjustment to Army life, and 3-month attrition. Moreover, the results are consistent with both theoretical descriptions of these scales and previous research (Ingerick et al., 2009; Knapp & Heffner, 2010). In some cases, the magnitudes of the correlations were smaller than what had been found in previous experimental research, however, and the TAPAS composite scores predicted key criteria at a lower rate. Nonetheless, because of the substantive differences between the research and IOT&E contexts, and the preliminary nature of the data, we cannot yet draw a definitive conclusion concerning the reasons for the differences between these settings. Several new scales (e.g., Generosity and Adjustment) showed statistically significant correlations with criteria, suggesting that future work should consider updating or revising the selection-oriented composites to enhance the validity of this tool.

With regard to classification potential, the results of the RMSD values on the mean differences for the overall TAPAS were comparatively smaller than those observed in the ASVAB. The magnitude of the differences varied by TAPAS scale, however, often in ways that are consistent with a theoretical understanding of the scale and the MOS. For example, the means for Physical Conditioning were higher for more physically-oriented MOS, such as 11B and 31B. The mean for the Intellectual Efficiency scale was highest for 68W, the most cognitively-oriented MOS in the sample. Additionally, the overall pattern of RMSD validity results suggests that TAPAS scores evidence differential prediction (or validity) that could enhance new Soldier classification over the ASVAB.

Taken together, these early evaluation results suggest that the TAPAS holds promise for both selection and classification-oriented purposes. Many of the scale-level coefficients are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. However, given the restricted nature of the matched criterion sample, these results should be considered highly preliminary. Future analyses should expand on these results by examining operational applications of the TAPAS, such as developing new selection and classification composites and determining the effect of various cut scores.

The second set of TOPS evaluation analyses will be conducted early in CY2011 based on data collected through December 2010. The sample sizes for this next evaluation are expected to be considerably larger, thus supporting additional analyses (e.g., re-examination of the will-do and can-do TAPAS composite scores) and yielding more generalizable results.

Utilization and Dissemination of Findings:

The research findings will be used by the U.S. Army Accessions Command, U.S. Army Recruiting Command, Army G-1, and Training and Doctrine Command to evaluate the effectiveness of tools used for Army applicant selection and assignment. With each successive set of findings, the Tier One Performance Screen can be revised and refined to meet Army needs and requirements.

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: EARLY RESULTS

CONTENTS

Page

CHAPTER 1: INTRODUCTION 1
Deirdre J. Knapp (HumRRO), Tonia S. Heffner and Len White (ARI)
Background 1
The Tier One Performance Screen (TOPS) 2
Evaluating TOPS 3
Overview of Report 4

CHAPTER 2: DATABASE DEVELOPMENT 5
D. Matthew Trippe, Laura Ford, Karen Moriarty, and Yuqui A. Cheng (HumRRO)
Description of Database and Sample Construction 6
Summary 9

CHAPTER 3: DESCRIPTION OF THE TOPS IOT&E PREDICTOR MEASURES 10
Stephen Stark, O. Sasha Chernyshenko, Fritz Drasgow (Drasgow Consulting Group), and Matthew T. Allen (HumRRO)
Tailored Adaptive Personality Assessment System (TAPAS) 10
TAPAS Background 10
Three Current Versions of TAPAS 11
TAPAS Scoring 13
TAPAS Initial Validation Effort 15
Initial TAPAS Composites 16
ASVAB Content, Structure, and Scoring 16
Summary 17

CHAPTER 4: PSYCHOMETRIC EVALUATION OF THE TAPAS 19
Matthew T. Allen, Michael J. Ingerick, and Justin A. DeSimone (HumRRO)
Empirical Comparison of the Three TAPAS Versions 19
Comparison of the TAPAS-95s with the TOPS IOT&E TAPAS 22
Summary 30

CHAPTER 5: DESCRIPTION AND PSYCHOMETRIC PROPERTIES OF CRITERION MEASURES 31
Karen O. Moriarty and Yuqui A. Cheng (HumRRO)
Training Criterion Measure Descriptions 32
Job Knowledge Tests (JKTs) 32
Performance Rating Scales (PRS) 32
Army Life Questionnaire (ALQ) 33
Administrative Criteria 35

CONTENTS (continued)

Page

Training Criterion Measure Scores and Associated Psychometric Properties 35
Job Knowledge Tests (JKTs) 35
Performance Rating Scales (PRS) 36
Army Life Questionnaire (ALQ) 38
Administrative Criterion Data 38
Summary 39

CHAPTER 6: INITIAL EVIDENCE FOR THE PREDICTIVE VALIDITY AND CLASSIFICATION POTENTIAL OF THE TAPAS 40

D. Matthew Trippe, Joseph P. Caramagno, Matthew T. Allen, and Michael J. Ingerick (HumRRO)

Predictive Validity 40
Analyses 40
Criterion-Related Validity Evidence 42
Classification Potential 45
Analyses 45
Cross-MOS Differences in TAPAS Score Profiles 46
Cross-MOS Differences in Predictive Validity Estimates 48
Summary and Conclusion 52

CHAPTER 7: SUMMARY AND A LOOK AHEAD 54

Deirdre J. Knapp (HumRRO), Tonia S. Heffner and Leonard A. White (ARI)

Summary of the TOPS IOT&E Method 54
Summary of Initial Evaluation Results 54
TAPAS Construct Validity 54
Validity for Soldier Selection 55
Potential for Soldier Classification 55
A Look Ahead 56

REFERENCES 57

APPENDIX A: BIVARIATE TAPAS CORRELATION TABLES A-1

APPENDIX B: COMPLETE TAPAS SUBGROUP MEAN DIFFERENCES B-1

APPENDIX C DESCRIPTIVE STATISTICS FOR THE FULL SCHOOLHOUSE SAMPLE C-1

APPENDIX D: SUPPLEMENTAL VALIDITY AND CLASSIFICATION TABLES D-1

Page

List of Tables

Table 2.1. Full TOPS Database Records by Relevant Characteristics 7
Table 2.2. Distribution of MOS in the Full Schoolhouse Database.....7
Table 2.3. Background and Demographic Characteristics of the TOPS Samples 8

Table 3.1. TAPAS Dimensions Assessed 12
Table 3.2. Descriptive Statistics for the ASVAB Based on the TOPS IOT&E Analysis Samples 18

Table 4.1. Standardized Mean Score and Standard Deviation Differences between TOPS IOT&E TAPAS Versions by Scale 20
Table 4.2. Standardized Differences in Scale Score Intercorrelations between the TOPS IOT&E TAPAS Versions by Dimension 22
Table 4.3. Standardized Mean Score and Standard Deviation Differences between EEEM TAPAS-95s and the TOPS IOT&E TAPAS by Version and Scale.....26
Table 4.4. Standardized Differences in Scale Score Intercorrelations between the EEEM TAPAS-95s and the TOPS IOT&E TAPAS by Version and Dimension.....27
Table 4.5. Differences in Scale Score Correlations between the TAPAS-95s and the TOPS IOT&E TAPAS with Individual Difference Variables 29

Table 5.1. Summary of Training Criterion Measures 31
Table 5.2. Example Training Performance Rating Scales 32
Table 5.3. ALQ Scales 34
Table 5.4. Descriptive Statistics and Reliability Estimates for Training Job Knowledge Tests (JKTs) in the Applicant Sample 36
Table 5.5. Descriptive Statistics and Reliability Estimates for Training Performance Rating Scales (PRS) in the Applicant Sample 37
Table 5.6. Descriptive Statistics and Reliability Estimates for the ALQ in the Applicant Sample.....38
Table 5.7. Descriptive Statistics for Administrative Criteria Based on the Applicant Sample.....39

Table 6.1. Incremental Validity Estimates for the TAPAS Scales over the AFQT for Predicting Select Performance- and Retention-Related Criteria 42
Table 6.2. Bivariate and Semi-Partial Correlations between the TAPAS Scales and Selected Criteria.....44

CONTENTS (continued)

Page

Table 6.3. Correlations between TAPAS Composite Scores and Select Performance and Retention-Related Criteria 45
Table 6.4. Average Root Mean Squared Differences in Mean TAPAS Scale Score Profiles for the Eight Target MOS 47
Table 6.5. Average Root Mean Squared Differences in Mean TAPAS Scale Score Profiles for the Expanded Sample of MOS 49
Table 6.6. Average Root Mean Squared Differences in Predictive Validity Estimates for Five Target MOS 51

Table A.1. TAPAS Intercorrelations for the 13-Dimension Computer-Adaptive (13D-CAT) Version (Applicant Sample) A-1
Table A.2. TAPAS Intercorrelations for the 15-Dimension Static (15D-Static) Version (Applicant Sample) A-2
Table A.3. TAPAS Intercorrelations for the 15-Dimension Computer-Adaptive (15D-CAT) Version (Applicant Sample) A-3
Table A.4. TAPAS-95s Intercorrelations from the Expanded Enlistment Eligibility Metrics (EEEM) Research A-4
Table A.5. TAPAS Intercorrelations for the 13-Dimension Computer-Adaptive (13D-CAT) Version (Accession Sample) A-4
Table A.6. TAPAS Intercorrelations for the 15-Dimension Static (15D-Static) Version (Accession Sample) A-5
Table A.7. TAPAS Intercorrelations for the 15-Dimension Computer-Adaptive (15D-CAT) Version (Accession Sample) A-5

Table B.1. TOPS Subgroup Mean Differences for Applicant Sample B-1
Table B.2. TOPS Subgroup Mean Differences for Accession Sample B-2

Table C.1. Descriptive Statistics for Training Criteria Based on the Full Schoolhouse Sample C-1
Table C.2. Descriptive Statistics for Schoolhouse Criteria by MOS (Full Schoolhouse Sample) C-3
Table C.3 Interrater Reliability Estimates for the Army-Wide and MOS-Specific PRS using the Full Schoolhouse Sample C-4
Table C.4. Army Life Questionnaire (ALQ) Intercorrelations for the Full Schoolhouse Sample C-4
Table C.5. MOS Job Knowledge Test (JKT) Correlations with the WTBD JKT in Full Schoolhouse Sample C-5

CONTENTS (continued)

Page

Table C.6. Army-Wide and MOS-Specific Performance Rating Scale (PRS) Intercorrelations for the Full Schoolhouse Sample C-5
Table C.7 Correlations between the Army Life Questionnaire (ALQ) and Job Knowledge Test (JKT) Scores for the Full Schoolhouse Sample C-6
Table C.8. Correlations between the Army Life Questionnaire (ALQ) and Performance Rating Scales (PRS) Scores for the Full Schoolhouse Sample C-7
Table C.9. Correlations between Job Knowledge Test (JKT) and Performance Rating Scale (PRS) Scores for the Full Schoolhouse Sample C-8
Table C.10. Descriptive Statistics for Administrative Criteria Based on the Applicant Sample by MOS C-9

Table D.1. Incremental Validity Estimates for the TAPAS Scales over the AFQT for Predicting Performance- and Retention-related Criteria D-1
Table D.2. Bivariate and Semi-Partial Correlations between the TAPAS Scales and Can-do Performance-related Criteria D-2
Table D.3. Bivariate and Semi-partial Correlations between the TAPAS Scales and Will-do Performance-related Criteria D-3
Table D.4. Bivariate and Semi-partial Correlations between the TAPAS Scales and Retention-related Criteria D-4
Table D.5. Correlations between TAPAS Can-do Composite Scores and Performance- and Retention-related Criteria D-5
Table D.6. Correlations between TAPAS Will-do Composite Scores and Performance- and Retention-related Criteria D-6
Table D.7. Mean TAPAS Scores for the Target and Expanded Sample of MOS D-7

List of Figures

Figure 1.1. TOPS Initial Operational Test & Evaluation (IOT&E). 3

Figure 2.1. Summary of TOPS schoolhouse (IMT) data sources. 5

Figure 2.2. Overview of TOPS database and sample generation process. 6

Figure 5.1. Relative overall performance rating scale. 33

TIER ONE PERFORMANCE SCREEN INITIAL OPERATIONAL TEST AND EVALUATION: EARLY ANALYSES

CHAPTER 1: INTRODUCTION

Deirdre J. Knapp (HumRRO), Tonia S. Heffner and Len White (ARI)

Background

The Personnel Assessment Research Unit (PARU) of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) is responsible for conducting manpower and personnel research for the Army. The focus of PARU's research is maximizing the potential of the individual Soldier through maximally effective selection, classification, and retention strategies.

In addition to educational, physical, and moral screens, the U.S. Army relies on a composite score from the Armed Services Vocational Aptitude Battery (ASVAB), the Armed Forces Qualification Test (AFQT), to select new Soldiers into the Army. Although the AFQT has proven to be and will continue to serve as a useful metric for selecting new Soldiers, other personal attributes, in particular non-cognitive attributes (e.g., temperament, interests, and values), are important to entry-level Soldier performance and retention (e.g., Knapp & Tremble, 2007).

In December 2006, the Department of Defense (DoD) ASVAB review panel—a panel of experts in the measurement of human characteristics and performance— released their recommendations (Drasgow, Embretson, Kyllonen, & Schmitt, 2006). Several of these recommendations focused on supplementing the ASVAB with additional measures for use in selection and classification decisions. The ASVAB review panel further recommended that the use of these measures be validated against performance criteria.

Just prior to release of the ASVAB review panel's findings, ARI initiated a longitudinal research effort, *Validating Future Force Performance Measures (Army Class),* to examine the prediction potential of several non-cognitive measures (e.g., temperament and person-environment fit) for Army outcomes (e.g., performance, attitudes, attrition). The Army Class research project is a 6-year effort that is being conducted with contract support from the Human Resources Research Organization (HumRRO; Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009). Experimental predictors were administered to new Soldiers in 2007 and early 2008. Since then, Army Class researchers have obtained attrition data from Army records and collected training criterion data on a subset of the Soldier sample. Job performance criterion data were collected from Soldiers in the Army Class longitudinal validation sample in 2009 (Knapp, Owens, & Allen, 2010) and a second round of job performance data is being collected in 2010-2011.

After the Army Class research was underway, ARI initiated the *Expanded Enlistment Eligibility Metrics (EEEM)* project (Knapp & Heffner, 2010). The EEEM goals were similar to Army Class, but the focus was specifically on Soldier selection (not classification) and the time horizon was much shorter. Specifically, EEEM required selection of one or more promising new predictor measures for immediate implementation. The EEEM project capitalized on the existing Army Class data collection procedure and, thus, the EEEM sample was a subset of the Army Class sample.

As a result of the EEEM findings, Army policy-makers approved an initial operational test and evaluation (IOT&E) of the *Tier One Performance Screen (TOPS)*. This report presents early analyses from the IOT&E of TOPS.

The Tier One Performance Screen (TOPS)

Six experimental pre-enlistment measures were included in the EEEM research (Allen, Cheng, Putka, Hunter, & White, 2010).[2] The "best bet" measures recommended to the Army for implementation were identified based on the following considerations:

- Incremental validity over AFQT for predicting important performance and retention-related outcomes
- Minimal subgroup differences
- Potential susceptibility to response distortion (e.g., faking good)
- Administration time requirements

The Tailored Adaptive Personality Assessment System (TAPAS; Stark, Chernyshenko, & Drasgow, 2010b) surfaced as the top choice, with the Work Preferences Assessment (WPA; Putka & Van Iddekinge, 2007) identified as another good option that was substantively different from the TAPAS. Specifically, TAPAS is a measure of personality characteristics (e.g., achievement, sociability) that capitalizes on the latest in testing technology whereas the WPA asks respondents to indicate their preference for various kinds of work activities and environments (e.g., "A job that requires me to teach others," "A job that requires me to work outdoors").

In May 2009, the Military Entrance Processing Command (MEPCOM) began administering TAPAS on the computer adaptive platform for the ASVAB (CAT-ASVAB). Initially TAPAS was to be administered only to Education Tier 1 (primarily high school diploma graduates), non-prior service applicants. The limitation to Education Tier 1 applicants was removed several months after the start so the Army could evaluate TAPAS across all types of applicants. The TAPAS administration by MEPCOM will continue through the fall of 2012.

The Tier One Performance Screen (TOPS) is intended to use non-cognitive measures to identify Education Tier 1 applicants who would likely perform differently (higher or lower) than would be predicted by their ASVAB scores. As part of the TOPS IOT&E, TAPAS scores are being used to screen out a small number of AFQT Category IV applicants.[3] Although the WPA is part of the TOPS IOT&E, it will not be considered for enlistment eligibility. The WPA is being prepared for MEPS administration with an expected administration start date of spring 2011.

Although the initial conceptualization for the IOT&E was to use TAPAS as a tool for "screening in" Education Tier 1 applicants with lower AFQT scores,[4] the economic conditions spurred a reconceptualization to a system that screens out low motivated applications with low AFQT scores. It is likely that the selection model in a fully operational system would adjust to

[2] These included several temperament measures, a situational judgment test, and two person-environment fit measures based on values and interests.
[3] Screening will expand to include a small number of Category IIIB applicants in Jul 2011.
[4] Initial supporting data analysis work focused on Category IIIB applicants (Allen et al., 2010), but TOPS currently targets those in Category IV.

fit with the changing applicant market. For example, at the present time, few applicants are being screened out based on TAPAS scores, not just because the passing scores are set quite low, but also because there are very few Category IV applicants being considered for enlistment due to the overwhelming availability of applicants in higher AFQT categories. Because many factors may impact how TAPAS would be used in the applicant screening process, TAPAS is administered to all Education Tier 1 and many Tier 2 non-prior service applicants who take the ASVAB in the MEPS.

Evaluating TOPS

Figure 1.1 illustrates the TOPS IOT&E research plan. To evaluate the non-cognitive measures (TAPAS and WPA), the Army is collecting training criterion data on Soldiers in eight target MOS[5] as they complete initial military training (IMT). The criterion measures include job knowledge tests (JKTs); an attitudinal person-environment fit assessment, the Army Life Questionnaire (ALQ); and performance rating scales (PRS) completed by the Soldiers' cadre. These measures are administered via the Internet at the schools for each of the eight target MOS. The process is overseen by Army personnel with guidance and support from both ARI and HumRRO. Course grades and completion rates are obtained from administrative records for all Soldiers who take the TAPAS, regardless of MOS.

Two waves of in-unit job performance data collection are also planned, both of which will attempt to capture Soldiers from across all MOS who completed the TAPAS (and WPA) during the application process. These measures again will include JKTs, the ALQ, and supervisor ratings. Finally, the separation status of all Soldiers who took the TAPAS is being tracked throughout the course of the research.

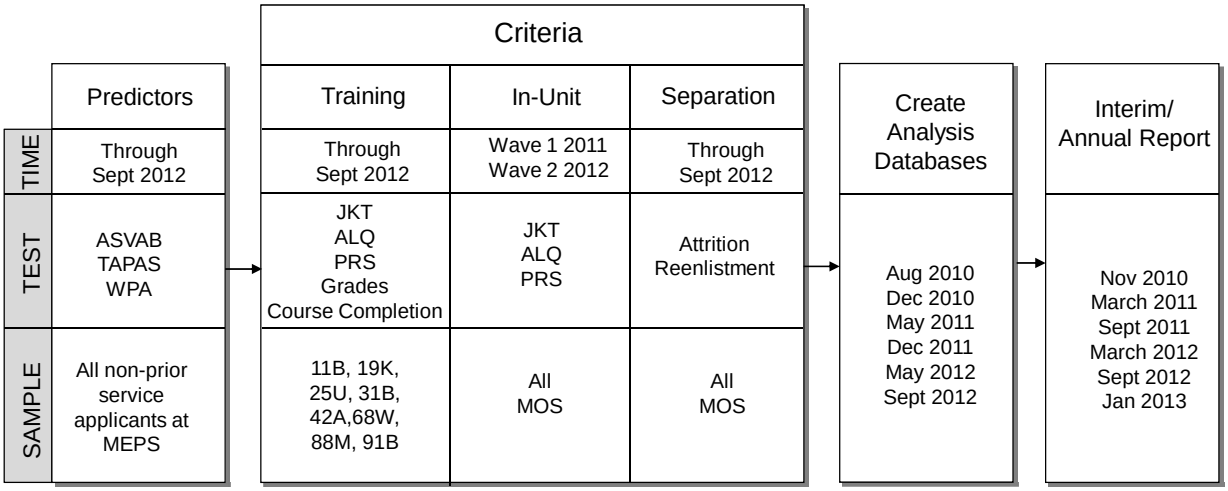

Figure 1.1. TOPS Initial Operational Test & Evaluation (IOT&E).

[5] The target MOS are Infantryman (11B), Armor Crewman (19K), Signal Support Specialist (25U), Military Police (31B), Human Resources Specialist (42A), Health Care Specialist (68W), Motor Transport Operator (88M), and Light Wheel Vehicle Mechanic (91B).

This report describes the initial effort to develop a criterion-related validation database and conduct evaluation analyses using data collected early in the TOPS IOT&E initiative. Additional analysis datasets and validation analyses will be prepared and conducted at 6-month intervals throughout the 3-year IOT&E period.

Overview of Report

Chapter 2 explains how the evaluation analysis databases are constructed, then describes characteristics of the samples resulting from construction of the first database in August 2010. Chapter 3 describes the TAPAS and ASVAB, including content and scoring. Chapter 4 offers an evaluation of TAPAS' psychometric characteristics. Chapter 5 describes the criterion measures included in this first analysis database, including their psychometric characteristics. Criterion-related validity analyses are presented in Chapter 6. The report concludes with Chapter 7, which summarizes this first attempt to evaluate TOPS and looks toward plans for future iterations of these evaluations.

CHAPTER 2: DATABASE DEVELOPMENT

D. Matthew Trippe, Laura Ford, Karen Moriarty, and Yuqui A. Cheng (HumRRO)

The Tier One Performance Screen (TOPS) database is assembled from a number of sources. In general, the database comprises predictor and criterion data obtained from administrative[6] and initial military training (IMT; or "schoolhouse") sources.

Schoolhouse records comprise assessment data collected from Soldiers and cadre at the locations identified in Figure 2.1. The outcome measures for the target MOS were specifically designed for this research and are not available from administrative sources. For the Soldiers, these assessments include job knowledge tests of Warrior Tasks and Battle Drills, MOS-specific tests, and a performance and attitudes questionnaire. For the cadre, the assessments are performance ratings scales on which they rate their Soldiers on Army-wide and MOS-specific performance dimensions.

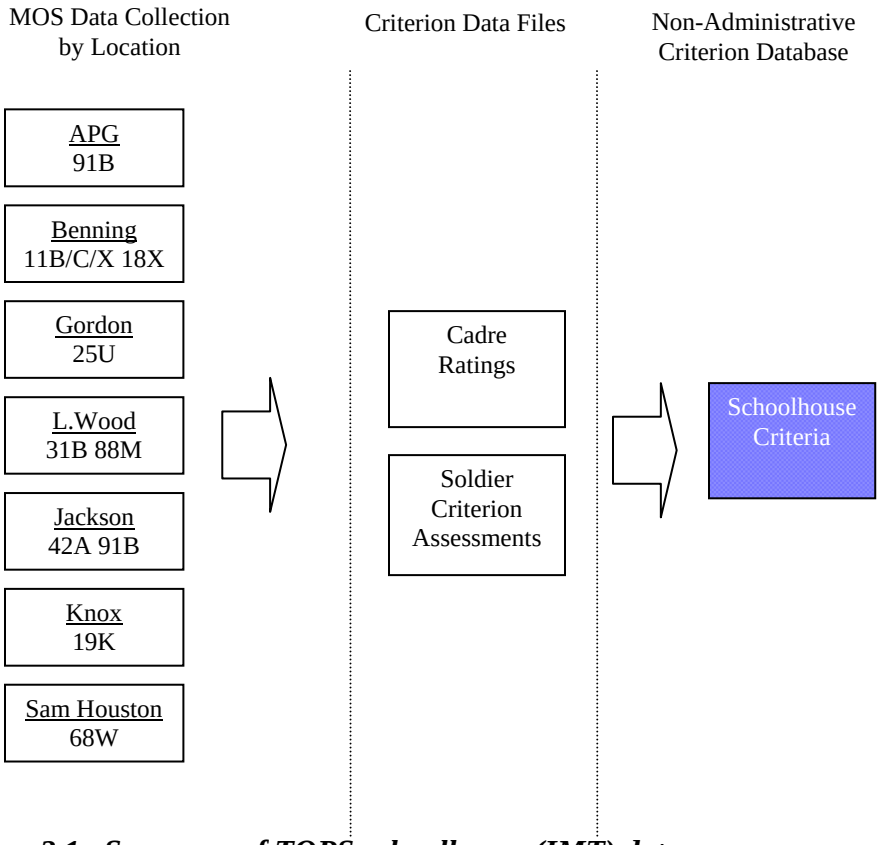

Figure 2.1. Summary of TOPS schoolhouse (IMT) data sources.

[6] Administrative data are collected from the following sources: (a) Military Entrance Processing Command (MEPCOM), (b) Army Human Resources Command (AHRC), (c) U.S. Army Accessions Command (USAAC), and (d) Army Training Support Center (ATSC).

More specific details regarding the composition of the analysis databases are conveyed in Figure 2.2. The white boxes within the figure represent database files, and shaded boxes represent samples on which descriptive or inferential analyses are conducted. Samples are formed by applying filters to a database such that it includes the observations of interest. The leftmost column in the figure summarizes the predictor data sources used to derive the two analysis samples (i.e., the "applicant" and "accession" samples). The middle column of the figure summarizes the criterion data sources, including IMT data from which the schoolhouse criterion sample is derived. Predictor and criterion data are merged to form the TOPS criterion-related analysis database (rightmost column).

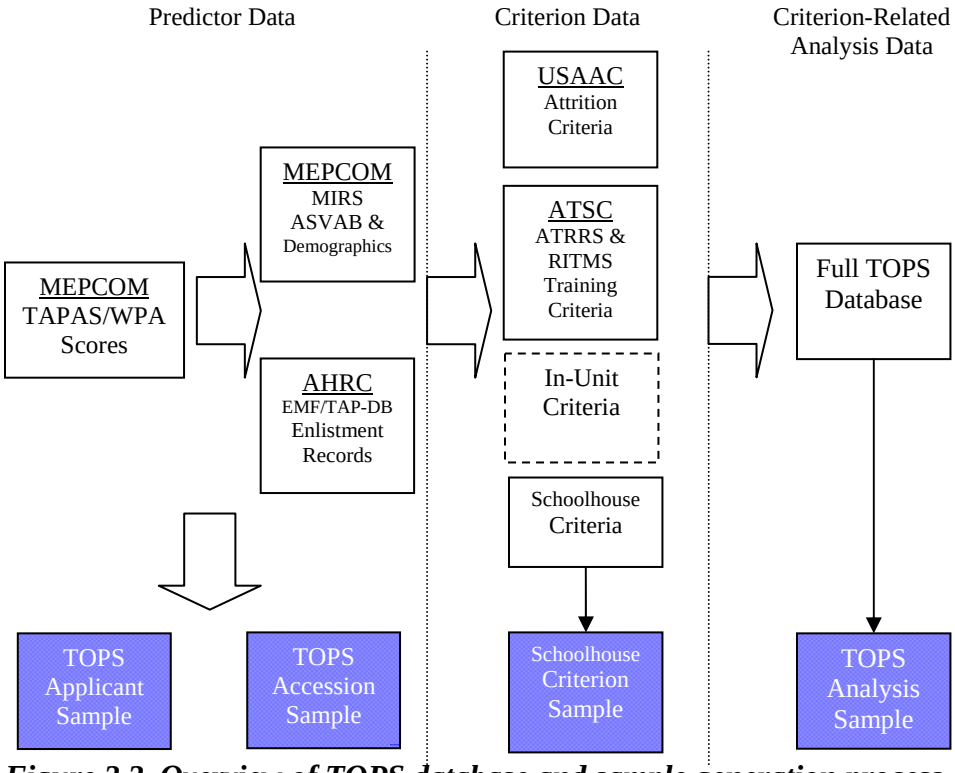

Figure 2.2. Overview of TOPS database and sample generation process.

Description of Database and Sample Construction

Table 2.1 summarizes the total sample contained in the August 2010 TOPS database by key variables that were used to create the samples on which analyses were conducted. The total sample includes all applicants regardless of whether they did or did not sign a contract. The majority of individuals in the database are classified as Education Tier 1, non-prior service, and AFQT Category I to IV (i.e., AFQT score $\geq$ 10). All analyses are restricted to these individuals, which results in elimination of approximately 11% of the total records in the database.

Table 2.1. Full TOPS Database Records by Relevant Characteristics

Variables	*N*	% of Total Sample (*N* = 60,485)
Education Tier		
Tier 1	56,548	93.5
Tier 2	2,189	3.6
Tier 3	1,748	2.9
Prior Service		
Yes	1,202	2.0
No or Missing	59,283	98.0
AFQT Category		
I	4,867	8.1
II	18,891	31.2
IIIA	11,809	19.5
IIIB	14,420	23.8
IV	9,446	15.6
V	1,052	1.7
Contract Status		
Signed	25,127	41.5
Not signed (as of Aug 10)	35,358	58.5
Total Tier 1, Non-prior service (NPS), AFQT ≥ 10[a]	53,964	89.2
Total Tier 1, NPS, AFQT ≥ 10, Contract signed[b]	24,177	40.0

[a]Constitutes the applicant sample.
[b]Constitutes the accession sample.

The number and percentage of each MOS represented in the schoolhouse criterion database is found in Table 2.2. The schoolhouse database comprises mainly 11B and 68W Soldiers. Other MOS represent 0.2% to 12% of the sample.

Table 2.2. Distribution of MOS in the Full Schoolhouse Database

MOS	Schoolhouse Criterion Database	
	n	*%*
11B/11C/11X/18X[a]	3,829	48.3
19K	12	0.2
25U	438	5.5
31B	465	5.9
42A	234	3.0
68W	1,744	22.0
88M	954	12.0
91B	246	3.1
Unknown	10	0.1
Total	7,932	100.0

[a]Soldiers in these MOS all participate in the same IMT course.

A detailed breakout of background and demographic characteristics observed in the analytic samples appears in Table 2.3. Regular Army Soldiers comprise a majority of the cases in each sample. AFQT categories follow an expected distribution. The samples are predominantly male, Caucasian, and non-Hispanic; however a significant percentage of Soldiers declined to provide information on race or ethnicity. The applicant sample was defined by limiting records in the full database to those who are non-prior service, Education Tier 1, and achieve an AFQT

score of at least 10. The accession sample was defined by further limiting the applicant sample to those Soldiers who signed an enlistment contract with the Army.

Table 2.3. Background and Demographic Characteristics of the TOPS Samples

	Applicant [a] *N* = 53,964		Accession [b] *N* = 24,177		Validation [c] *N* = 3,592		Schoolhouse Validation *N* = 397	
Characteristic	*n*	%	*n*	%	*n*	%	*n*	%
Component								
Regular	32,728	60.7	18,495	76.5	2,839	79.0	239	60.2
ARNG	14,323	26.5	2,086	8.6	518	14.4	117	29.5
USAR	6913	12.8	3,596	14.9	235	6.5	41	10.3
MOS								
11B/11C/11X/18X	2271	4.2	1,360	5.6	782	21.8	188	47.3
19K	166	0.3	134	0.6	73	2.0	1	.3
25U	299	0.6	164	0.7	34	1.0	7	1.8
31B	933	1.7	416	1.7	112	3.1	39	9.8
42A	426	0.8	313	1.3	61	1.7	25	6.3
68W	1172	2.2	844	3.5	222	6.2	57	14.4
88M	1207	2.2	777	3.2	188	5.2	63	15.9
91B	809	1.5	548	2.3	100	2.8	17	4.3
Other	10,247	19.0	7,584	31.4	1,877	52.3	--	--
Unknown	36,434	67.5	12,037	49.8	143	4.0	--	--
AFQT Category								
I	4543	8.4	2,066	8.6	343	9.6	27	6.8
II	17,447	32.3	8,687	35.9	1,337	37.2	148	37.3
IIIA	10,752	19.9	5,557	23.0	850	23.7	93	23.4
IIIB	12,877	23.9	6,688	27.7	914	25.5	106	26.7
IV	8345	15.5	1,179	4.9	148	4.1	23	5.8
Gender								
Female	10,491	19.4	3,935	16.3	494	13.8	46	11.6
Male	43,473	80.6	20,242	83.7	3,098	86.3	351	88.4
Race								
African American	5,871	10.9	2,152	8.9	268	7.5	30	7.6
American Indian	394	0.7	176	0.7	23	0.6	1	.3
Asian	1,142	2.1	499	2.1	56	1.6	6	1.5
Caucasian	35,298	65.4	15,913	65.8	2,240	62.4	246	62.0
Other	735	1.4	348	1.4	98	2.7	13	3.2
Decline to Answer	10,524	19.5	5,089	21.1	907	25.3	101	25.4
Ethnicity								
Hispanic/Latino	7224	13.4	2,964	12.3	246	6.9	23	5.8
Not Hispanic	36,250	67.2	16,369	67.7	2,483	69.1	274	69.0
Decline to Answer	10,490	19.4	4,844	20.0	863	24.0	100	25.2

[a] Sample limited to Soldiers who had no prior service, Education Tier 1, and AFQT ≥ 10.
[b] The accession sample includes those in the applicant sample further limited to Soldiers who signed a contract.
[c] The validation sample includes those in the accession sample further limited to Soldiers who had at least one criterion variable.

The accession sample amounts to roughly half of the applicant sample. This reduction is likely due in part to the lack of maturity of some administrative records, which may not yet reflect the true accession status for all records. The validation sample described in Table 2.3 includes 3,592 Soldiers. Those included in the validation sample are Soldiers that meet all of the inclusion criteria for the accession sample and also have at least one criterion variable that is used in the validity or classification analyses reported in Chapters 6 and 7. However, the number of Soldiers included in any individual validity or classification analysis is generally much smaller. The exact number of Soldiers included in a given analysis depends on the criterion

variable involved. Specific sample details on each criterion variable are provided in the subsequent analysis chapters. Generally speaking, 3-month attrition data accounts for approximately 2,800 of these records and the approximately 700 administrative graduation and exam records represent the next most available criterion data source. Although there were 7,932 Soldiers in the full schoolhouse database, only 438 Soldiers had taken the TAPAS when they applied for enlistment. This disconnect was due largely to the delayed entry of many Soldiers. That is, we believe that most of the Soldiers tested at the schools had taken their pre-enlistment tests before MEPCOM started administering the TAPAS to applicants. The problem was exacerbated by the gradual introduction of the TAPAS across MEPS locations so that early in the IOT&E, not all MEPS were yet actively participating. We expect that future analysis databases will show a far higher match between Soldiers tested in the schools and those tested pre-enlistment.

Summary

The TOPS data was assembled by merging TAPAS scores, administrative records, and IMT data into one master database. The TAPAS and IMT data were both rigorously cleaned in preparation for scoring. A total of 60,485 applicants took the TAPAS, 53,964 of which were in the applicant sample primarily used for analysis. The applicant sample was determined by excluding Education Tier 2, AFQT Category V, and prior service applicants from the master database. However, of that 53,964, only 3,592 (6.7%) had a criterion variable record, and only 397 (0.7%) had valid IMT data. Because of this low match rate, the analyses reported in the remainder of this report should be treated as highly preliminary.

CHAPTER 3: DESCRIPTION OF THE TOPS IOT&E PREDICTOR MEASURES

Stephen Stark, O. Sasha Chernyshenko, Fritz Drasgow (Drasgow Consulting Group), and Matthew T. Allen (HumRRO)

The purpose of this chapter is to describe the predictor measures investigated in the initial months of the TOPS IOT&E. The central predictor under investigation in this analysis is the Tailored Adaptive Personality Assessment System (TAPAS; Stark, Chernyshenko, & Drasgow, 2010b), while the baseline predictor used by the Army is the ASVAB. We begin this chapter by describing the TAPAS, including previous research and scoring methodology. This is followed by a brief description of the versions administered as part of the TOPS IOT&E. We finish by briefly describing the ASVAB and its psychometric properties.

Tailored Adaptive Personality Assessment System (TAPAS)

TAPAS Background

TAPAS is a new personality measurement tool developed by Drasgow Consulting Group (DCG) under the Army's Small Business Innovation Research (SBIR) program. The system builds on the foundational work of the Assessment of Individual Motivation (AIM; White & Young, 1998) by incorporating features designed to promote resistance to faking and by measuring narrow personality constructs (i.e., facets) that are known to predict outcomes in work settings. Because TAPAS uses item response theory (IRT) methods to construct and score items, it can be administered in multiple formats: (a) as a fixed length, *nonadaptive test* where examinees respond to the same sequence of items or (b) as an *adaptive* test where each examinee responds to a unique sequence of items selected to maximize measurement accuracy for that specific examinee.

TAPAS uses a recently developed IRT model for multidimensional pairwise preference items (MUPP; Stark, Chernyshenko, & Drasgow, 2005) as the basis for constructing, administering, and scoring personality tests that are designed to reduce response distortion (i.e., faking) and yield normative scores even with tests of high dimensionality (Stark, Chernyshenko, & Drasgow 2010a). TAPAS items consist of pairs of personality statements for which a respondent's task is to choose the statement in each pair that is "more like me." The two statements composing each item are matched in terms of social desirability and often represent different dimensions. As a result, respondents have a difficult time discerning which answers improve their chances of being enlistment eligible. Because they are less likely to know which dimensions are being used for selection, they are less likely to discern which statements measure those dimensions, and they are less likely to keep track of their answers on several dimensions simultaneously so as to provide consistent patterns of responses across the whole test. Without knowing which answers impact their eligibility status, respondents should not be able to increase their scores on selection dimensions as easily as when traditional, single statement measures are used.

The use of a formal IRT model also greatly increases the flexibility of the assessment process. A variety of test versions can be constructed to measure personality dimensions that are relevant to specific work contexts, and the measures can be administered via paper-and-pencil or

computerized formats. If test design specifications are comparable across versions, the respective scores can be readily compared because the metric of the statement parameters has already been established by calibrating response data obtained from a base or reference group (e.g., Army recruits). The same principle applies to adaptive testing, wherein each examinee receives a different set of items chosen specifically to reduce the error in his or her trait scores at points throughout the exam. Adaptive item selection enhances test security because there is less overlap across examinees in terms of the items presented. Even with constraints governing the repetition and similarity of the psychometric properties of the statements composing TAPAS items, we estimate that over 100,000 possible pairwise preference items can be crafted from the current 15-dimension TAPAS pool.

Another important feature of TAPAS is that it contains personality statements representing 22 narrow personality traits. The TAPAS trait taxonomy was developed using the results of several large scale factor-analytic studies with the goal of identifying a comprehensive set of non-redundant narrow traits. These narrow traits, if necessary or desired, can be combined to form either the Big Five (the most common organization scheme for narrow personality traits) or any other number of broader traits (e.g., Integrity or Positive Core Self-Evaluations). This is advantageous for applied purposes because TAPAS versions can be created to fit a wide range of applications and are not limited to a particular service branch or criterion. Selection of specific TAPAS dimensions can be guided by consulting results of an unpublished meta-analytic study performed by DCG that mapped the 22 TAPAS dimensions to several important organizational criteria for military and civilian jobs (e.g., task proficiency, training performance, attrition).

Three Current Versions of TAPAS

As part of the TOPS IOT&E, three versions of the TAPAS were administered. The first version was a 13-dimension computerized adaptive test (CAT) containing 104 pairwise preference items. This version is referred to as the TAPAS-13D-CAT. TAPAS-13D-CAT was administered from May 4, 2009 to July 10, 2009 to over 2,200 Army and Air Force recruits.[7] In July 2010, ARI decided to expand the TAPAS to 15 dimensions by adding the facets of Adjustment from the Emotional Stability domain and Self-Control from the Conscientiousness domain. Test length was also increased to 120 items. Two 15-dimension TAPAS tests were created. One version was nonadaptive (static), so all examinees answered the same sequence of items; the other was adaptive, so each examinee answered items tailored to his or her trait level estimates. The TAPAS-15D-Static was administered from mid-July to mid-September of 2009 to all examinees, and later to smaller numbers of examinees at some MEPS. The adaptive version, referred to as TAPAS-15D-CAT, was introduced in September and Army and Air Force recruits continue to be administered this version of TAPAS. Table 3.1 shows the facets assessed by the 13-dimension and 15-dimension measures. Descriptive statistics for the TAPAS are provided in Chapter 4, along with analyses examining comparability across versions.

[7] Note that MEPCOM also is administering the TAPAS to Air Force applicants on an experimental basis.

Table 3.1. TAPAS Dimensions Assessed

Facet Name	Brief Description	"Big Five" Broad Factor
Dominance	High scoring individuals are domineering, "take charge" and are often referred to by their peers as "natural leaders."	Extraversion
Sociability	High scoring individuals tend to seek out and initiate social interactions.	
Attention Seeking	High scoring individuals tend to engage in behaviors that attract social attention; they are loud, loquacious, entertaining, and even boastful.	
Generosity	High scoring individuals are generous with their time and resources.	Agreeableness
Cooperation	High scoring individuals are trusting, cordial, non-critical, and easy to get along with.	
Achievement	High scoring individuals are seen as hard working, ambitious, confident, and resourceful.	Conscientiousness
Order	High scoring individuals tend to organize tasks and activities and desire to maintain neat and clean surroundings.	
Self Control[a]	High scoring individuals tend to be cautious, levelheaded, able to delay gratification, and patient.	
Non-Delinquency	High scoring individuals tend to comply with rules, customs, norms, and expectations, and they tend not to challenge authority.	
Adjustment[a]	High scoring individuals are worry free, and handle stress well; low scoring individuals are generally high strung, self-conscious and apprehensive.	Emotional Stability
Even Tempered	High scoring individuals tend to be calm and stable. They don't often exhibit anger, hostility, or aggression.	
Optimism	High scoring individuals have a positive outlook on life and tend to experience joy and a sense of well-being.	
Intellectual Efficiency	High scoring individuals are able to process information quickly and would be described by others as knowledgeable, astute, and intellectual.	Openness To Experience
Tolerance	High scoring individuals scoring are interested in other cultures and opinions that may differ from their own. They are willing to adapt to novel environments and situations.	
Physical Conditioning	High scoring individuals routinely participate in vigorous sports or exercise and enjoy physical work.	Other

[a]Not included in TAPAS-13D-CAT.

TAPAS Scoring

TAPAS scoring is based on the MUPP IRT model originally proposed by Stark (2002). The model assumes that when person *j* encounters stimuli *s* and *t* (which, in our case, correspond to two personality statements), the person considers whether to endorse *s* and, independently, considers whether to endorse *t*. This process of independently considering the two stimuli continues until one and only one stimulus is endorsed. A preference judgment can then be represented by the joint outcome (Agree with *s*, Disagree with *t*) or (Disagree with *s*, Agree with *t*). Using a 1 to indicate agreement and a 0 to indicate disagreement, the outcome (1,0) indicates that statement *s* was endorsed but statement *t* was not, leading to the decision that *s* was preferred to statement *t*; an outcome of (0,1) similarly indicates that stimulus *t* was preferred to *s*. Thus, the probability of endorsing a stimulus *s* over a stimulus *t* can be formally written as

$$P_{(s>t)_i}(\theta_{d_s},\theta_{d_t}) = \frac{P_{st}\{1,0\,|\,\theta_{d_s},\theta_{d_t}\}}{P_{st}\{1,0\,|\,\theta_{d_s},\theta_{d_t}\} + P_{st}\{0,1\,|\,\theta_{d_s},\theta_{d_t}\}},$$

where:

$P_{(s>t)_i}(\theta_{d_s},\theta_{d_t}) =$ probability of a respondent preferring statement *s* to statement *t* in item *i*,

$i =$ index for items (i.e., pairings), where $i = 1$ to I,

$d =$ index for dimensions, where $d = 1, \ldots, D$, d_s represents the dimension assessed by statement *s*, and d_t represents the dimension assessed by statement t,

$s, t =$ indices for first and second statements, respectively, in an item,

$(\theta_{d_s},\theta_{d_t}) =$ latent trait scores for the respondent on dimensions d_s and d_t respectively,

$P_{st}(1,0\,|\,\theta_{d_s},\theta_{d_t}) =$ joint probability of endorsing stimulus *s* and not endorsing stimulus *t* given latent trait scores $(\theta_{d_s},\theta_{d_t})$,

and

$P_{st}(0,1\,|\,\theta_{d_s},\theta_{d_t}) =$ joint probability of not endorsing stimulus *s* and endorsing stimulus *t* given latent trait scores $(\theta_{d_s},\theta_{d_t})$.

With the assumption that the two statements are evaluated independently, and with the usual IRT assumption that only θ_{d_s} influences responses to statements on dimension d_s and only θ_{d_t} influences responses to dimension d_t (i.e., local independence), we have

$$P_{(s>t)_i}(\theta_{d_s},\theta_{d_t}) = \frac{P_s(1\,|\,\theta_{d_s})P_t(0\,|\,\theta_{d_t})}{P_s(1\,|\,\theta_{d_s})P_t(0\,|\,\theta_{d_t}) + P_s(0\,|\,\theta_{d_s})P_t(1\,|\,\theta_{d_t})},$$

where

$P_s(1 \mid \theta_{d_s}), P_s(0 \mid \theta_{d_s})$ = probability of endorsing/not endorsing stimulus *s* given the latent trait value θ_{d_s},

and

$P_t(0 \mid \theta_{d_t}), P_t(1 \mid \theta_{d_t})$ = probability of endorsing/not endorsing stimulus *t* given latent trait θ_{d_t}.

The probability of preferring a particular statement in a pair thus depends on θ_{d_s} and θ_{d_t}, as well as the model chosen to characterize the process for responding to the individual statements. Toward that end, Stark (2002) proposed using the dichotomous case of the generalized graded unfolding model (GGUM; Roberts, Donoghue, & Laughlin, 2000), which has been shown to fit personality data reasonably well (Chernyshenko, Stark, Drasgow, & Roberts, 2007; Stark, Chernyshenko, Drasgow, & Williams, 2006).

Test scoring is done via Bayes modal estimation. For a vector of latent trait values, $\tilde{\theta} = (\theta_{d'=1}, \theta_{d'=2}, ..., \theta_{d'=D})$, this involves maximizing:

$$L(\tilde{u}, \tilde{\theta}) = \left\{ \prod_{i=1}^{n} \left[P_{(s>t)_i} \right]^{u_i} \left[1 - P_{(s>t)_i} \right]^{1-u_i} \right\} * f(\tilde{\theta}) \quad ,$$

where $\tilde{u}$ is a binary response pattern, $P_{(s>t)}$ is the probability of preferring statement *s* to statement *t* in item *i*, and $f(\tilde{\theta})$ is a *D*-dimensional prior density distribution, which, for simplicity, is assumed to be the product of independent normals, $\prod_{d'=1}^{D} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\theta_{d'}^2}{2\sigma^2}}$.

Taking the natural log, for convenience, the above equation can be rewritten as:

$$\ln L(\tilde{u}, \tilde{\theta}) = \sum_{i=1}^{n} \left[(u_i) \ln P_{(s>t)_i} + (1 - u_i) \ln(1 - P_{(s>t)_i}) \right] + \sum_{d'=1}^{D} \left[\ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \frac{\theta_{d'}^2}{2\sigma^2} \right] \quad ,$$

leaving the following set of equations to be solved numerically:

$$\frac{\partial \ln L}{\partial \tilde{\theta}} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta_{d'=1}} \\ \frac{\partial \ln L}{\partial \theta_{d'=2}} \\ \vdots \\ \frac{\partial \ln L}{\partial \theta_{d'=D}} \end{bmatrix} = 0$$

This equation can be solved numerically to obtain a vector of trait score estimates for each respondent via a *D*-dimensional maximization procedure (e.g., Press, Flannery, Teukolsky, & Vetterling, 1990), involving the posterior and its first derivatives. Standard errors for TAPAS trait scores are estimated using a replication method developed by Stark and colleagues (2010a). In brief, this method involves using the IRT parameter estimates for the items that were administered to generate 30 new response patterns based on an examinee's TAPAS trait scores. The resulting simulated response patterns are then scored and the standard deviations of the respective trait estimates over the 30 replications are used as standard errors for the original TAPAS values. In a recent simulation study (Stark, Chernyshenko, & Drasgow, 2010c), this new replication method provided standard error estimates that were much closer to the empirical (true) standard deviations than previously used approaches (i.e., based on the approximated inverse Hessian matrix or a jack-knife approach).

TAPAS Initial Validation Effort

Initial predictive and construct-related validity evidence on the TAPAS was collected during ARI's *Expanded Enlistment Eligibility Metrics* (EEEM) research project in 2007-2009 (Knapp & Heffner, 2010). As described in Chapter 1, the EEEM effort was conducted in conjunction with ARI's *Army Class* longitudinal validation of multiple experimental non-cognitive predictor measures. In the EEEM project, new Soldiers completed a 12-dimension, 95-item nonadpative (or static) version of TAPAS, called TAPAS-95s. TAPAS-95s was administered as a paper questionnaire that included an information sheet showing respondents a sample item and illustrating how to properly record their answers to the "questions" that followed. Respondents were specifically instructed to choose the statement in each pair that was "more like me" and that they must make a choice even if they found it difficult to do so. Item responses were coded dichotomously and scored using an updated version of Stark's (2002) computer program for MUPP trait estimation.

Overall, the TAPAS-95s showed evidence of construct and criterion validity. Intellectual Efficiency and Curiosity, for example, showed moderate positive correlations with AFQT and correlations of .35 with each other. This was expected, given that both facets tap the intellectance aspects of the Big Five factor, Openness to Experience. The same two traits exhibited similarly positive, but smaller, correlations with Tolerance, another facet of Openness reflecting comfortableness around others having different customs, values, or beliefs (Chernyshenko, Stark, Woo, & Conz, 2008). TAPAS-95s dimensions also showed incremental validity over AFQT in predicting several performance criteria. For example, when TAPAS trait scores were added to the regression analysis based on a sample of several hundred Soldiers, the multiple correlation increased by .35 for the prediction of physical fitness, .20 for the prediction of disciplinary incidents, and .11 for the prediction of 6-month attrition. None of these criteria were predicted well by AFQT alone (predictive validity estimates were consistently below .10).

In sum, the EEEM research showed TAPAS-95s to be a viable assessment tool with the potential to enhance new Soldier selection. Trait scores exhibited construct validity evidence with respect to other measures and criterion-related validity estimates were fairly high for outcomes not predicted well by AFQT. Based on the results of this research and taking into consideration the unique advantages of TAPAS (e.g., flexibility and resistance to faking), the Army chose to test the measure in an applicant environment.

Initial TAPAS Composites

In addition to the validation analyses described above, an initial Education Tier 1 performance screen was developed from the TAPAS-95s scales for the purpose of testing in an applicant setting (Allen, Cheng, Putka, Hunter, & White, 2010). This was accomplished by (a) identifying key criteria of most interest to the Army, (b) sorting these criteria into "can-do" and "will-do" categories, and (c) selecting composite scales corresponding to the can-do and will-do criteria, taking into account both theoretical rationale and empirical results. The result of this process was two composite scores.

1. Can-Do Composite: The TOPS can-do composite consists of five TAPAS scales and is designed to predict can-do criteria such as MOS-specific job knowledge, AIT exam grades, and graduation from AIT/OSUT.

2. Will-Do Composite: The TOPS will-do composite consists of five TAPAS scales (three of which overlap with the can-do composite) and is designed to predict will-do criteria such as physical fitness, adjustment to Army life, effort, and support for peers.

The target population for these composites was AFQT Category IIIB applicants, though, due to changing recruitment priorities (as described in Chapter 1) the target group was later changed to AFQT Category IV applicants. Initial validity and subgroup difference results suggest that cut scores based on these two composites were promising for selecting applicants with high potential and with minimal subgroup differences.

ASVAB Content, Structure, and Scoring

The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple aptitude battery of nine tests administered by the Military Entrance Processing Command. Most military applicants take the computer adaptive version of ASVAB (i.e., the CAT-ASVAB). Scores on the ASVAB tests are combined to create composite scores for use in (a) selecting applicants into the Army and (b) classifying them to an MOS. The Armed Forces Qualification Test (AFQT) comprises the Verbal Expression[8] (VE), Arithmetic Reasoning (AR), and Math Knowledge (MK) tests (AFQT = 2*VE + AR + MK). Applicants must meet a minimum AFQT score to be eligible to serve in the military and the Services favor high-scoring applicants for enlistment (e.g., through enlistment bonuses). For classification, scores on the ASVAB tests are combined to form nine Aptitude Area (AA) composites.[9] An applicant must receive a minimum score on the MOS-relevant AA composite(s) to qualify for classification to that MOS. For example, applicants must score a 95 in both the Electronics (EL) and Signal Communications (SC) AA composites to qualify as a Signal Support Specialist (25U).

Descriptive statistics for the AFQT, ASVAB tests, and AA composites are reported in Table 3.2 for the two main analysis samples described in Chapter 2 (i.e., the Applicant and Accession samples). The AFQT mean for the Accession Sample is slightly higher than the mean

[8] Verbal Expression is a scaled combination of the Word Knowledge (WK) and Paragraph Comprehension (PC) tests.
[9] A tenth AA composite, General Technical (GT), is not used for enlisted Army selection or classification and therefore is not included here.

found in previous research on a similar population (EEEM; Knapp & Heffner, 2010; *M* = 57.28 versus 61.61 in the TOPS sample), suggesting this sample may have higher general cognitive aptitude than previous samples. The AFQT standard deviation for the TOPS sample, however, is slightly larger than in previous research (EEEM *SD* = 20.15; TOPS *SD* = 20.72).

Summary

The purpose of this chapter was to describe the predictor measures used as part of the TOPS IOT&E. Three versions of the experimental measure—the TAPAS—were administered as part of the TOPS IOT&E. The TAPAS is unique from typical personality measures because it uses forced-choice pairwise items and IRT to promote resistance to faking. Initial validation research conducted as part of EEEM was promising enough to warrant an IOT&E. Both the individual TAPAS scales and can-do and will-do composites formed as part of EEEM are evaluated in subsequent chapters. The baseline instrument was the ASVAB, which consists of multiple tests that are formed into selection (i.e., AFQT) and classification (i.e., AA) composites. Results suggest that the AFQT mean and standard deviation were higher in the TOPS Accession sample than in the EEEM research, suggesting the present sample may have higher general cognitive aptitude than previous samples.

Table 3.2. Descriptive Statistics for the ASVAB Based on the TOPS IOT&E Analysis Samples

	Applicant Sample					Accession Sample				
Measure/Scale	*n*	*M*	*SD*	*Min*	*Max*	*n*	*M*	*SD*	*Min*	*Max*
AFQT	53,964	57.69	23.71	10	99	24,177	61.61	20.72	10	99
ASVAB Subtests										
General Science (GS)	39,229	51.74	8.61	21	76	18,977	52.96	7.92	24	76
Arithmetic Reasoning (AR)	39,229	52.58	7.96	22	72	18,977	53.74	7.15	28	72
Word Knowledge (WK)	39,229	51.32	8.40	20	76	18,977	52.48	7.63	20	76
Paragraph Comprehension (PC)	39,229	52.75	7.39	25	69	18,977	53.84	6.75	25	69
Math Knowledge (MK)	39,229	53.21	7.24	26	73	18,977	54.19	6.54	28	73
Electronics Information (EI)	39,228	52.41	9.32	16	84	18,976	53.59	8.82	16	84
Auto and Shop Information (AS)	39,228	50.94	9.68	23	86	18,976	51.89	9.39	24	86
Mechanical Comprehension (MC)	39,227	53.78	8.67	23	82	18,976	55.03	8.08	23	82
Assembling Objects (AO)	39,034	55.12	7.96	25	70	18,871	56.06	7.53	26	70
ASVAB Aptitude Area Composites										
Clerical (CL)	39,227	105.85	14.70	59.86	151.45	18,976	108.41	12.78	67.96	150.33
Combat (CO)	39,227	106.15	15.61	54.38	159.85	18,976	108.86	13.84	68.24	158.15
Electronics (EL)	39,227	105.91	15.64	55.78	159.59	18,976	108.63	13.82	68.18	156.91
Field Artillery (FA)	39,227	106.27	15.53	54.34	159.14	18,976	108.98	13.74	70.08	156.44
General Maintenance (GM)	39,227	105.79	16.05	54.88	160.64	18,976	108.52	14.34	65.91	159.64
Mechanical Maintenance (MM)	39,227	105.33	17.03	55.94	163.37	18,976	108.00	15.58	59.43	163.37
Operators and Food Service (OF)	39,227	105.79	16.04	56.74	159.88	18,976	108.52	14.28	68.60	157.65
Signal Communication (SC)	39,227	106.15	15.27	54.37	158.52	18,976	108.84	13.42	67.70	154.98
Skill Technical (ST)	39,227	105.98	15.28	56.86	156.85	18,976	108.68	13.39	68.88	152.61

Note. Applicant Sample = Non-prior service, Education Tier 1, AFQT Category IV and above. Accession Sample = Non-prior service, Education Tier 1, AFQT Category IV and above, signed contract.

CHAPTER 4: PSYCHOMETRIC EVALUATION OF THE TAPAS

Matthew T. Allen, Michael J. Ingerick, and Justin A. DeSimone (HumRRO)

The purpose of this chapter is to conduct a psychometric evaluation of the TAPAS in an applicant setting.[10] Specifically, we begin by comparing the psychometric characteristics (means, standard deviations, and intercorrelations) of the three versions of the TAPAS to one another. This is followed by an empirical comparison of the TOPS versions of the TAPAS with the TAPAS-95s, which was administered as part of the EEEM research (see Chapter 1).

Empirical Comparison of the Three TAPAS Versions

As described in Chapter 3, three versions of the TAPAS were administered as part of the TOPS research: (a) a computer-adaptive 13-dimension version (13D-CAT), (b) a static 15-dimension version (15D-Static), and (c) a computer-adaptive 15-dimension version (15D-CAT). Although the three versions were intended to be comparable, they should not be seen as parallel. All versions were based on the same statement pool, but the dimensionality, test length, and/or design specifications (i.e., the blueprints) varied. To determine whether the three versions were sufficiently equivalent to treat as one measure in subsequent analyses, we compared the three versions based on the (a) mean dimension scores and standard deviations and (b) intercorrelations among the dimension scores. The means and standard deviations of the raw dimension scores for the three TAPAS versions are summarized in Table 4.1. To compare the magnitude of the mean differences, standardized mean differences (i.e., Cohen's *d*) were computed for each TAPAS scale using the following formula:

$$d = M_{GROUP1} - M_{GROUP2} / SD_{POOLED} \quad (1)$$

Cohen's (1988) rule of thumb suggests that 0.20 to 0.30 should be considered a small effect, 0.50 a medium effect, and 0.80 or above a large effect. The differences between standard deviations were compared with an *F*-test, which was computed with the following formula:

$$F = SD^2_{GROUP1} / SD^2_{GROUP2} \quad (2)$$

We did not compute statistical significance tests for either the mean or standard deviation differences due to the large sample sizes of the three groups. Because of the large sample sizes, even small differences would be considered statistically significant using traditional null hypothesis testing. Accordingly, we focused on the effect size estimates when comparing the three versions.

[10] Although operational in the strictest sense of the term, the TOPS IOT&E applies a low screen to such few applicants (those in Education Tier 1 scoring in AFQT Category IV), that applicants may recognize that the scores are unlikely to matter for them personally. Thus, we use the term "applicant" rather than "operational" setting.

Table 4.1. Standardized Mean Score and Standard Deviation Differences between TOPS IOT&E TAPAS Versions by Scale

	TOPS TAPAS Version											
	13D-CAT (n = 1,311)		15D-Static (n = 8,224)		15D-CAT (n = 42,130)		Cohen's d			F-test		
Composite/Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	$d_{13D\text{-}15DS}$	$d_{13D\text{-}15C}$	$d_{15DS\text{-}15C}$	$F_{15DS\text{-}13D}$	$F_{13D\text{-}15C}$	$F_{15DS\text{-}15C}$
By Individual Composite/Scale												
Achievement	.234	0.493	.275	0.503	.150	0.480	-0.08	0.18	0.26	0.96	1.05	1.09
Adjustment	--	--	.159	0.582	-.005	0.570	--	--	0.29	--	--	1.04
Attention Seeking	-.224	0.557	-.246	0.528	-.194	0.533	0.04	-0.06	-0.10	1.11	1.09	0.98
Cooperation	.027	0.390	-.070	0.392	-.061	0.375	0.25	0.23	-0.03	0.99	1.08	1.10
Dominance	.072	0.600	-.026	0.589	.035	0.591	0.17	0.06	-0.10	1.04	1.03	1.00
Even Tempered	.126	0.514	.253	0.480	.159	0.477	-0.26	-0.07	0.20	1.15	1.16	1.01
Generosity	-.172	0.426	-.196	0.449	-.203	0.430	0.05	0.07	0.02	0.90	0.98	1.09
Intellectual Efficiency	.099	0.608	-.086	0.593	-.018	0.587	0.31	0.20	-0.12	1.05	1.07	1.02
Non-Delinquency	.105	0.457	.117	0.457	.088	0.459	-0.03	0.04	0.06	1.00	0.99	0.99
Optimism	.175	0.464	.261	0.511	.134	0.462	-0.17	0.09	0.27	0.83	1.01	1.22
Order	-.416	0.568	-.397	0.575	-.431	0.548	-0.03	0.03	0.06	0.98	1.08	1.10
Physical Conditioning	-.019	0.617	-.048	0.619	.026	0.629	0.05	-0.07	-0.12	1.00	0.96	0.97
Self-Control	--	--	.098	0.527	.058	0.532	--	--	0.07	--	--	0.98
Sociability	-.026	0.622	-.209	0.594	-.037	0.594	0.31	0.02	-0.29	1.09	1.09	1.00
Tolerance	-.240	0.598	-.249	0.588	-.231	0.570	0.02	-0.02	-0.03	1.03	1.10	1.06
Can-Do Composite	.739	1.406	.821	1.382	.513	1.373	-0.06	0.16	0.22	1.03	1.05	1.01
Will-Do Composite	.669	1.319	.844	1.225	.616	1.247	-0.14	0.04	0.18	1.16	1.12	0.96
Averages	*M*	*SD*	*M*	*SD*	*M*	*SD*	$\lvert d \rvert$	$\lvert d \rvert$	$\lvert d \rvert$	*F*	*F*	*F*
All TAPAS Scales	-.020	0.532	-.024	0.532	-.035	0.522	0.12	0.08	0.13	1.01	1.05	1.04

Note. Results are limited to the "Applicant Sample" (Non-prior service, Education Tier 1, AFQT Category IV and above). 13D = TAPAS 13D-CAT, 15DS = TAPAS 15D-Static, 15C = TAPAS 15D-CAT.

The results in Table 4.1 suggest that despite the differences in length, dimensionality, and design specifications acknowledged at the outset, the three versions of the TAPAS were quite similar in terms of their means and standard deviations. The *d* statistics ranged from a low of 0.02 to a high of 0.31 and the average absolute values of the *d* statistics were all below 0.20, which is considered a "small" difference. Only 7 of the 47 pairwise comparisons were above 0.25, or a quarter of a standard deviation. Three of these differences were between the 13D-CAT and 15D-Static, and four were between the 15D-CAT and 15D-Static. The 13D-CAT and 15D-CAT versions had the most similar means. Overall, this suggests that the number of dimensions (13 or 15) and format (static or adaptive) of the TAPAS had little effect on the facet mean and standard deviation scores, though the format led to slightly more differences. The largest differences tended to be for the Sociability, Intellectual Efficiency, Optimism, and Cooperation scales. In terms of standard deviations, all of the *F*-values were near 1.0, suggesting that the variances are roughly equivalent across the three versions. The one exception to this pattern was the Optimism scale, which exhibited an *F* value of 1.22 between the 15D-Static and 15D-CAT versions of the TAPAS.

Another basis for examining the consistency between the different TAPAS versions is in the pattern of intercorrelations among the dimension scores. For example, if Dominance is positively correlated with Achievement in one version of the TAPAS, we would reasonably expect a positive correlation of a similar magnitude to be found in another version of the TAPAS, regardless of any mean score differences between the versions. Specifically, we would expect a similar pattern of intercorrelations among the dimensions that are theoretically or taxonomically related, such as the facets underlying the Big Five (see Table 3.1, Chapter 3). To test the similarity of the intercorrelation matrices for the three versions, we computed a Standardized Root Mean Square Residual (SRMR). Following Hu and Bentler (1999), the SRMR was computed using the following formula,

$$\text{SRMR} = \sqrt{\left\{ 2 \sum_{i=1}^{p} \sum_{j=1}^{i} [(s_{ij} - \hat{\sigma}_{ij}) / (s_{ii}\, s_{jj})]^2 \right\} / p(p+1)} \qquad (3)$$

where s_{ij} is the observed covariances for one group (i.e., applicants completing one TAPAS version), σ_{ij} is the observed covariances for the comparison group, s_{ii} and s_{jj} are the observed standard deviations, and p is the number of observed variables. SRMR is a commonly used fit index in confirmatory factor analysis. Following Hu and Bentler's (1999) recommendations, we interpreted SRMRs that are close to zero as very similar, while those above .08 are interpreted as different.

The results of the SRMR analysis can be found in Table 4.2, while the full correlation matrices are in Appendix A. We report the SRMRs comparing (a) the full correlation matrices, (b) the matrices corresponding to the Big Five, and (c) the matrices corresponding to the can-do and will-do TAPAS composites. The SRMRs based on (b) and (c) were computed to better diagnose where the systematic differences, if any, were among the versions that may otherwise be lost from an examination of the full matrices. Overall, the results suggest that the patterns of intercorrelations were very similar between the three TAPAS versions. No SRMR values were above .08, and only one SRMR value – comparing the 13D-CAT and 15D-Static versions – was above .05. Further examination of the bivariate correlations between the two versions (see Appendix A) suggests that the main sources of discrepancy were on the Achievement, Cooperation, and Even Tempered scales.

For example, the Achievement/Cooperation ($r_{13D\text{-}CAT}$ = .07, $r_{15D\text{-}Static}$ = -.03; Z = 3.26, p < .01) and Achievement/Optimism ($r_{13D\text{-}CAT}$ = .13, $r_{15D\text{-}Static}$ = .26; Z = -4.33, p < .01) correlations were significantly different between the two versions. Overall however, the results of the SRMR analysis suggest the patterns of intercorrelations for the two versions are quite similar.

Table 4.2. Standardized Differences in Scale Score Intercorrelations between the TOPS IOT&E TAPAS Versions by Dimension

Composite/ Scale Score Profile	$SRMR_{13C\text{-}15DS}$	$SRMR_{13C\text{-}15C}$	$SRMR_{15DS\text{-}15C}$
All TAPAS Scales	.0574	.0357	.0468
By Big Five Factor			
Agreeableness	.0059	.0019	.0040
Conscientiousness	.0243	.0243	.0246
Emotional Stability	.0060	.0178	.0370
Extraversion	.0348	.0259	.0182
Openness to Experience	.0169	.0061	.0107
By TOPS Composite			
Can-Do	.0480	.0208	.0395
Will-Do	.0475	.0280	.0276

Note. 13D-CAT, *n* = 1,311. 15D-Static, *n* = 8,224. 15D-CAT, *n* = 42,130. Values reported are standardized root mean squared residuals (SRMR). SRMR values greater than .08 are bolded (Hu & Bentler, 1999). Results are limited to the Applicant Sample (non-prior service, Education Tier 1, AFQT Category IV and above). 13D = TAPAS 13D-CAT, 15DS = TAPAS 15D-Static, 15C = TAPAS 15D-CAT.

In summary, the results suggest that the means, standard deviations, and intercorrelations for the three versions are comparable. Therefore, it would be appropriate to combine scores from the three versions in the same analysis, provided that the scores are standardized within version to account for any small scaling differences.

Comparison of the TAPAS-95s with the TOPS IOT&E TAPAS

Previous work testing temperament measures such as the Assessment of Individual Motivation (AIM) under operational conditions has found high levels of socially desirable responding that lead to criterion-related validity coefficients approaching zero[11] (White, Young, Hunter, & Rumsey, 2008). These results motivated the continuation of research on fake-resistant personality measures and, in fact, led to the development of the TAPAS. For obvious reasons, the critical evidence concerning the effectiveness of TAPAS for selection applications lies in its performance under operational conditions, so comparisons of internal and relational properties across examinee groups taking the test in a common environment (e.g., military entrance processing stations) but under different instructions (operational vs. research only) would be highly valued. Because the data for such comparisons were not available for this report, an alternative was to compare means, intercorrelations, and validities for the three versions of TAPAS explored in the IOT&E to the TAPAS-95s administered in the EEEM research project (Knapp & Heffner, 2010). Despite the systematic differences in the examinee pools and test forms discussed previously, such analyses were seen as useful for providing at least a rough indication of the effect of situational factors on the test scores.

[11] Additional work with the AIM as a component of the Tier Two Accession Screen (TTAS) has demonstrated that it contributes to the prediction of attrition for applicants who are non-high school diploma graduates.

To address this issue, we conducted analyses similar to those from the previous section. Specifically, we compared TOPS TAPAS versions to the TAPAS-95s based on (a) the facet score means and standard deviations, (b) intercorrelations among facet scores, and (c) correlations between the dimension scores and external individual difference variables (e.g., demographics, AFQT scores). To ensure that the two samples were as comparable as possible, the results of the TOPS TAPAS analyses were limited to respondents that were Education Tier 1, non-prior service, AFQT Category IV and above, and had signed a contract with the Army (i.e., the Accession Sample described in Chapter 2). The results for the TAPAS-95s were also limited to Education Tier 1, non-prior service Soldiers.

It is important to note that the TOPS TAPAS versions and TAPAS-95s are not parallel measures because many statements used in the TAPAS-95s were also included in the TOPS TAPAS statement *pool*, but parameters for some statements were re-estimated in accordance with refinements to the TAPAS trait taxonomy. For example, statements from the TAPAS-95s "Optimism" facet were reallocated to the "Adjustment" and "Optimism" facets before the TOPS implementation. In addition, statement parameters for Tolerance, Order, Cooperation, and Even Tempered were revised based on additional data that were collected, thus making direct comparisons between TOPS and EEEM difficult.

In sum, substantive differences between the EEEM context and the present one are enumerated below.

1. The TAPAS-95s was administered via paper and pencil, while the three versions of the TOPS TAPAS were computer-administered.

2. The TAPAS-95s was static, while two of the three TOPS TAPAS versions were adaptive.

3. The TAPAS-95s assessed 12 dimensions using 95 items, whereas the TOPS TAPAS versions assessed 13 dimensions with 104 items or 15 dimensions with 120 items.

4. The TAPAS-95s was administered to Soldiers who had already accessed into the Army, whereas the TOPS TAPAS versions were administered to an applicant sample.

5. The TAPAS-95s was administered in an environment where the Army was having difficulty meeting its recruiting mission, whereas TOPS TAPAS was administered in a poor economic environment (McMichael, 2008; 2009; Schafer, 2007) in which recruiting was less challenging. As a result of these economic conditions, the Army became more selective in its recruiting and accessioning process during the course of the TOPS research.

Despite these aforementioned differences, substantial score inflations in operational settings and/or large changes in intercorrelations or correlations with external variables for the TOPS TAPAS versions could signal that the test is functioning differently as compared to research settings.

Table 4.3 presents the mean and standard deviations for 10 scales found in both TAPAS-95s and the three TOPS TAPAS versions. The scales with the smallest standardized mean differences were the Achievement (Avg. $|d| = 0.12$), Attention Seeking (Avg. $|d| = 0.13$), Non-Delinquency (Avg. $|d| = 0.02$), and Physical Conditioning (Avg. $|d| = 0.20$) scales. The Tolerance (Avg. $|d| = 0.27$) and Dominance (Avg. $|d| = 0.27$) scales also had standardized mean differences below 0.30. The Even Tempered (Avg. $|d| = 1.12$) and Order (Avg. $|d| = 0.70$) scales evidenced the largest mean differences, but these scores were based on parameters that were updated prior to TOPS, so the difference in means might be explained to some extent by changes in the IRT metrics. Also, certain facet scores such as Physical Conditioning decreased for the TOPS TAPAS as compared to TAPAS-95s, which would not be expected if faking were present.

The standard deviations of the TOPS TAPAS dimension scores reported in Table 4.3 were generally lower than the corresponding standard deviations observed on the TAPAS-95s. The average F values reflecting the difference in the standard deviations between the three TOPS TAPAS versions and the TAPAS-95s were consistently close to 2.00. With regard to scores on the individual dimensions, the Tolerance (Avg. $F = 1.31$), Intellectual Efficiency (Avg. $F = 1.21$), and Dominance (Avg. $F = 1.01$) standard deviations were most similar between the two settings, while the Cooperation (Avg. $F = 5.17$), Attention Seeking (Avg. $F = 2.22$), Even Tempered (Avg. $F = 2.51$), and Non-Delinquency (Avg. $F = 2.22$) scores demonstrated the largest differences. The magnitude and pattern of the differences in the standard deviations between the two settings were generally the same across the three TOPS TAPAS versions.

Another way to compare TAPAS-95s and TOPS TAPAS is to examine the consistency of their relationship with each other and with key individual difference variables. Correlations are useful because they are unaffected by linear transformations, associated with, for example, changing means or IRT recalibrations. Marked differences across settings or versions of a test could provide insights into how test construction practices affect item responding and ultimately construct and predictive validities.

With this in mind, we compared the patterns of intercorrelations among the facet scores from the TAPAS-95s to those observed in the TOPS TAPAS using the SRMR statistic described previously. The SRMR results are reported in Table 4.4, while the correlation matrices used to compute the SRMR are reported in Appendix A (Tables A.4–A.7). Note that we did not compute SRMRs for the Agreeableness and Emotional Stability dimensions because there was only one scale in each that was included in both the TOPS and EEEM studies. The Optimism scale was excluded from these analyses due to the content changes described above. Overall, we found few differences between the three TOPS TAPAS versions and the TAPAS-95s within the groupings where we would expect the most stable relationships (i.e., within Big Five dimension). The differences in matrices for the can-do composite were also relatively small. The larger differences in the two matrices were found for all of the TAPAS scales and the will-do composite.

The main source of difference in the intercorrelation matrices most often involved the Attention Seeking (Avg. Total $\Delta| r | = .10$) scale, which had the largest average difference between the three versions of the TOPS TAPAS and the TAPAS-95s. For example, the Attention Seeking scale had four intercorrelations where the average difference was above .10: (a) Cooperation (Avg. $\Delta| r | = .11$), (b) Intellectual Efficiency (Avg. $\Delta| r | = .13$), (c) Non-

Delinquency (Avg. Δ| *r* | = .23), and (d) Achievement (Avg. Δ| *r* | = .15). Other scales that had large average differences include Cooperation (Avg. Total Δ| *r* | = .09), and Dominance (Avg. Total Δ| *r* | = .08). The Order (Avg. Total Δ| *r* | = .03) and Physical Conditioning (Avg. Total Δ| *r* | = .05) scales had the smallest average differences.

Table 4.3. Standardized Mean Score and Standard Deviation Differences between EEEM TAPAS-95s and the TOPS IOT&E TAPAS by Version and Scale

	TAPAS Version													
	EEEM (95s) (n = 3,381)		13D-CAT (n = 786)				15D-Static (n = 4,258)				15D-CAT (n = 18,217)			
Scale	*M*	*SD*	*M*	*SD*	*d*	*F*	*M*	*SD*	*d*	*F*	*M*	*SD*	*d*	*F*
Achievement	.160	.625	.230	.495	0.12	1.59	.286	.512	0.22	1.49	.157	.481	-0.01	1.69
Attention Seeking	-.127	.797	-.206	.554	-0.10	2.07	-.236	.521	-0.17	2.35	-.192	.534	-0.11	2.23
Cooperation	-.282	.865	.027	.375	0.39	5.32	-.089	.390	0.30	4.92	-.048	.377	0.48	5.26
Dominance	-.144	.603	.070	.591	0.36	1.04	-.045	.608	0.16	0.98	.028	.600	0.29	1.01
Even Tempered	-.491	.764	.145	.497	0.88	2.36	.261	.479	1.21	2.55	.181	.473	1.27	2.61
Intellectual Efficiency	-.187	.647	.121	.596	0.48	1.18	-.046	.589	0.23	1.20	.011	.579	0.34	1.25
Non-Delinquency	.120	.661	.128	.430	0.01	2.37	.128	.448	0.01	2.18	.107	.455	-0.03	2.11
Order	-.034	.636	-.464	.560	-0.69	1.29	-.427	.574	-0.65	1.23	-.462	.551	-0.76	1.33
Physical Conditioning	.128	.712	.000	.609	-0.19	1.37	-.040	.627	-0.25	1.29	.033	.626	-0.15	1.29
Tolerance	-.420	.673	-.261	.599	0.24	1.26	-.259	.591	0.26	1.30	-.238	.575	0.31	1.37

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract).

Table 4.4. Standardized Differences in Scale Score Intercorrelations between the EEEM TAPAS-95s and the TOPS IOT&E TAPAS by Version and Dimension

	TAPAS Version		
Composite/ Scale Score Profile	13D-CAT (n = 786)	15D-Static (n = 4,258)	15D-CAT (n = 18,217)
All TAPAS Scales	.0754	**.0800**	**.0810**
By Big Five Factor			
Agreeableness	n/a	n/a	n/a
Conscientiousness	.0166	.0151	.0192
Emotional Stability	n/a	n/a	n/a
Extraversion	.0305	.0687	.0339
Openness to Experience	.0442	.0344	.0420
By TOPS Composite			
Can-Do	.0471	.0630	.0450
Will-Do	.0600	**.0984**	**.0867**

Note. Values reported are standardized root mean squared residuals (SRMR). SRMR values greater than .08 are bolded (Hu & Bentler, 1999). Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). The raw TAPAS scores were used in this analysis.

We also computed the correlations (or point-biserial correlation for binary demographic variables) between TAPAS dimension scores and four variables: (a) AFQT score, (b) race, (c) ethnicity, and (d) gender. For this analysis, the TOPS TAPAS versions were combined into one overall set of TAPAS scales by:

1. Filtering out participants that were not part of the sample of interest (i.e., those that were not in the "Applicant Sample" – Tier 1, non-prior service, AFQT Category IV or above), and

2. Standardizing the variables within version using a *z*-transformation, completed by subtracting each score from the mean for that version and dividing by the standard deviation.

This standardized version of the overall TAPAS was also used in the analyses described in Chapter 6. Once the correlations were computed, the TAPAS-95s and TOPS TAPAS results were compared using two statistics. The first was the squared difference between the correlations (Δr^2). The second was Fisher's Z test of the equality of two correlations, which can be expressed with the following formula (Cohen, Cohen, Aiken, & West, 2003):[12]

$$Z = \frac{z'_1 - z'_2}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}} \qquad (4)$$

[12] Note that Fisher's Z assumes that the variables under consideration are normally distributed. However, the dichotomous variables used in this analysis (race, ethnicity, and gender) are not normally distributed and, therefore, violate this assumption. Nevertheless, given that the purpose of this analysis was to measure the relative magnitude of the difference between two coefficients and that the Fisher's Z is appropriate for the AFQT/TAPAS correlations, the Fisher's Z was used for the dichotomous variables as well. However, this limitation should be kept in mind when interpreting these results.

where z'_1 and z'_2 are the logarithmic transformations of the correlations for groups 1 and 2 and n_1 and n_2 are the sample sizes. Values above 1.96 or less than -1.96 are considered statistically significant.

The results of this analysis are presented in Table 4.5. We generally found weak relations between TAPAS dimension scores and key individual difference variables. Very few of the correlations were above .10, and many were not statistically significant despite the large sample sizes. However, there were exceptions. For example, Intellectual Efficiency and the can-do TAPAS composite (which includes the Intellectual Efficiency scale) were strongly related to Solders' AFQT scores in both the EEEM TAPAS-95s and TOPS TAPAS versions. Second, Tolerance was positively correlated with all three demographic variables (race, ethnicity, and gender), suggesting that minority subgroups (Blacks, Hispanics, and females) tended to score higher on the Tolerance scale than the majority subgroups (Whites, Non-Hispanics, and males). The Generosity scale was also positively correlated with gender, suggesting that females score higher on that scale than males, while Physical Conditioning was negatively correlated with gender, suggesting that males score higher on that scale than females. While there were a number of other statistically significant correlations between the individual correlations and these demographic variables, the magnitude was generally small. This finding is further supported by the subgroup mean differences, presented as a reference in Appendix B.

There were differences between the TAPAS-95s and the TOPS TAPAS, as measured by the Δr^2 estimates, but the Δr^2 values were consistently .03 or less. Although a number of the Z comparisons were statistically significant, this was likely primarily due to the large sample sizes available for these analyses. The correlations demonstrating the largest differences between the two settings involved the Attention Seeking scale with AFQT, and the Dominance scale with gender. The Attention Seeking scale was negatively correlated with AFQT in EEEM, and positively correlated with AFQT in TOPS. Finally, the Dominance scale was positively correlated with gender in EEEM, and negatively correlated in TOPS. Despite these apparent differences, there was no systematic pattern of results to suggest that the relationship between these individual difference variables and the TAPAS changed fundamentally from one setting to the other.

Table 4.5. Differences in Scale Score Correlations between the TAPAS-95s and the TOPS IOT&E TAPAS with Individual Difference Variables

	EEEM				TOPS (Standardized)				Difference Metrics							
Scale	AFQT *r*	Race *r*	Eth *r*	Sex *r*	AFQT *r*	Race *r*	Eth *r*	Sex *r*	AFQT $(\Delta r)^2$	Race $(\Delta r)^2$	Eth $(\Delta r)^2$	Sex $(\Delta r)^2$	AFQT *Z*	Race *Z*	Eth *Z*	Sex *Z*
Achievement	**.06**	-.02	-.01	**.08**	**.07**	**-.02**	**-.02**	.00	.00	.00	.00	.01	-0.71	-0.28	0.70	**4.17**
Adjustment	.	.	.	.	**.09**	**-.04**	**-.06**	**-.11**	.	.	.	.	.	.	.	.
Attention Seeking	**-.07**	-.02	-.01	-.01	**.09**	**-.03**	**-.02**	**-.04**	.03	.00	.00	.00	**-8.97**	0.52	0.32	1.35
Cooperation	**-.04**	-.01	.00	**.05**	-.01	.01	.00	.00	.00	.00	.00	.00	-1.39	-1.05	0.00	**2.36**
Dominance	**.06**	**.08**	-.01	**.09**	**.07**	.00	**.02**	**-.05**	.00	.01	.00	.02	-0.90	**3.81**	-1.36	**7.48**
Even Tempered	**.14**	.02	.00	**-.04**	**.06**	.01	**-.02**	**-.03**	.01	.00	.00	.00	**3.96**	0.78	0.90	-0.63
Generosity	.	.	.	.	**-.07**	**.05**	**.02**	**.15**	.	.	.	.	.	.	.	.
Intellectual Efficiency	**.38**	.02	-.04	**-.07**	**.42**	**-.03**	**-.05**	**-.07**	.00	.00	.00	.00	-2.12	**2.91**	0.87	0.38
Non-Delinquency	**.06**	-.01	-.04	**.14**	.00	.01	**-.03**	**.05**	.00	.00	.00	.01	**2.83**	-1.27	-0.35	**4.57**
Optimism	.	.	.	.	.00	.00	.00	-.01	.	.	.	.	.	.	.	.
Order	**-.04**	**.06**	.02	**.11**	**-.15**	**.07**	**.05**	**.05**	.01	.00	.00	.00	**6.27**	-0.45	-1.37	**3.29**
Physical Conditioning	.00	.01	.01	**-.12**	**.03**	**-.06**	**-.03**	**-.14**	.00	.00	.00	.00	-1.52	**3.39**	1.79	1.18
Self-Control	.	.	.	.	.00	**.07**	**.03**	.01	.	.	.	.	.	.	.	.
Sociability	.	.	.	.	**-.09**	**-.02**	.00	.01	.	.	.	.	.	.	.	.
Tolerance	.02	**.12**	**.08**	**.10**	**-.01**	**.08**	**.10**	**.13**	.00	.00	.00	.00	1.80	1.84	-1.21	-1.67

Note. EEEM AFQT *n* = 3,362, EEEM Race *n* = 3,194, EEEM Ethnicity *n* = 2,833, EEEM Gender *n* = 3,368. TOPS AFQT *n* = 22,475-23,261, TOPS Race *n* = 16,909-17,416, TOPS Ethnicity *n* = 18,166-18,649, TOPS Gender *n* = 22,475-23,261. All of the demographic variables were coded as 1 or 0, with 1 being the minority subgroup: Race (1=Black, 0=White), Ethnicity (1=Hispanic, 0=Non-Hispanic), and Gender (1=Female, 0=Male). Δr^2 = the squared difference between the TOPS and EEEM TAPAS correlations. *Z* = The difference between the TOPS and EEEM TAPAS correlations as determined using Fisher's *Z* test. Values above 1.96 are bolded. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract).

Summary

To test whether the psychometric characteristics of the TAPAS were consistent across versions (13D-CAT, 15D-Static, 15D-CAT) and settings (EEEM vs. IOT&E), we conducted a number of diagnostic and comparative analyses. The results of these analyses suggest:

1. The three versions of the TAPAS (13D-CAT, 15D-Static, and 15D-CAT) were consistent with one another in terms of their means, standard deviations, and patterns of intercorrelations. The two computer-adaptive versions of the TAPAS were particularly similar. However, there were some mean differences for individual scales, suggesting the need to standardize within these three versions to account for scaling differences if versions other than the 15D-CAT are used in future assessments.

2. The standard deviations for the TOPS TAPAS were, on average, smaller than in the EEEM research, suggesting either (a) the TOPS population is narrower on these facets or (b) participants are responding in a way that is reducing the available variance for each scale.

3. Some of the TAPAS scales were more similar across the research and operational settings than others. For example, the psychometric properties for the Attention Seeking scale changed substantially from one setting to another, while the Tolerance and Physical Conditioning scales were similar across the two settings.

4. With a few exceptions, the TAPAS scales showed no bias as they were not strongly related to key individual difference variables (AFQT scores, race, ethnicity, and gender). Additionally, the patterns of these relationships were generally consistent from the EEEM to TOPS settings.

Keeping in mind that previous research has shown large differences between the experimental and operational use of temperament measures (White et al., 2008), these results suggest that the use of the TAPAS in an operational setting is promising. Although there were some differences in scale score means and standard deviations across the two settings, these differences could be explained by differences in test specifications and IRT metrics or other environmental factors rather than socially desirable responding.

CHAPTER 5: DESCRIPTION AND PSYCHOMETRIC PROPERTIES OF CRITERION MEASURES

Karen O. Moriarty and Yuqui A. Cheng (HumRRO)

Training criterion measures such as job knowledge tests (JKTs), performance rating scales (PRS), and attitudinal data captured on a self-report questionnaire were used to validate the TAPAS. These measures were originally developed for the training phase of the *Army Class* project (Moriarty, Campbell, Heffner, & Knapp., 2009), but modified, where needed, for inclusion in the TOPS IOT&E. As with *Army Class*, we used administrative data to expand the criterion space. Table 5.1 summarizes the training criterion measures.

Table 5.1. Summary of Training Criterion Measures

Criterion Measure	Description
Soldier/Cadre Reported	
Job Knowledge Tests (JKT)	MOS-specific JKTs measure Soldiers' knowledge of basic facts, principles, and procedures required of Soldiers in training for a particular MOS. Each JKT includes a mix of item formats (e.g., multiple-choice, multiple-response, and rank order). The Warrior Tasks and Battle Drills (WTBD) JKT measures knowledge that is general to all enlisted Army Soldiers.
Performance Rating Scales (PRS)	PRS measure Soldiers' training performance on two categories: (a) MOS-specific (e.g., learns preventive maintenance checks and services, learns to troubleshoot vehicle and equipment problems) and (b) Army-wide (e.g., exhibits effort, supports peers, demonstrates physical fitness). The PRS are completed by drill sergeants or training cadre.
Army Life Questionnaire (ALQ)	ALQ measures Soldiers' self-reported attitudes and experiences through IMT. The training ALQ focuses on Soldiers' attitudes and experiences in IMT and includes 13 scales that cover (a) Soldiers' commitment and retention-related attitudes, and (b) Soldiers' performance and adjustment.
Administrative	
Attrition	Attrition data were obtained on participating Regular Army Soldiers at 3 months time in service (TIS).
Initial Military Training (IMT) Criteria	These data provide information concerning how many Soldiers restarted IMT and for what reasons, and the number of times Soldiers restarted training.
AIT School Grades	Schoolhouse grades for Soldiers in Advanced Individual Training (AIT).

Training Criterion Measure Descriptions

Job Knowledge Tests (JKTs)

Depending upon the MOS, many JKT items were drawn from items originally developed in prior ARI projects (Campbell & Knapp, 2001; Collins, Le, & Schantz, 2005; Knapp & Campbell, 2006). Most of the JKT items are in a multiple-choice format with two to four response options. However, other formats, such as multiple response (i.e., check all that apply), rank ordering, and matching are also used. The items make use of visual images to make them more realistic and to reduce reading requirements for the test.

As noted, the JKTs were originally developed for the Army Class project. Prior to finalizing them for use in this project, the items were reviewed to ensure they were of high quality. First, we reviewed the comments Soldiers provided about the assessments during the Army Class testing sessions and made corrections where necessary. For example, several Soldiers did not know the meaning of the word, "demarcate," so we changed that word to "mark." Second, we reviewed item statistics from the Army Class data and dropped items that had poor item statistics (e.g., low item-total correlations). Finally, results of the Army Class JKT analyses suggested that the training JKTs were too difficult, so we eliminated the more difficult items to protect the content validity of the assessments.

Performance Rating Scales (PRS)

The PRS also have roots in previous research (see Moriarty et al., 2009 for details). Table 5.2 provides example scales. The number of dimensions per set of scales ranges from five to nine. The scales were completed by cadre members of the target Soldiers. The scales ranged from 1 (lowest) to 7 (highest) and included a "not observed" option for instances where the cadre did not have an opportunity to observe a Soldier's performance. They are in the format of a behaviorally-anchored rating scale (BARS), where raters provide one rating per dimension using several examples of high, medium, and low performance as anchors.

Table 5.2. Example Training Performance Rating Scales

MOS/AW	Name	Description
Army-Wide	Effort	Puts forth individual effort in study, practice, preparation, and participation activities to complete AIT/OSUT requirements to meet individual Soldier expectations.
MOS-Specific	Learns Safety Procedures	How well has the Soldier learned to follow safety procedures, being alert to possible dangerous or hazardous situations and taking steps to protect self, other Soldiers, and equipment?

For Army Class, the performance anchors were organized into high and low performance for the Army-wide scales; there were no medium performance anchors (Moriarty et al., 2009). For the TOPS project, we converted the bipolar statements into high, medium, and low anchors to be consistent with the MOS-specific PRS. We also wrote additional items where appropriate. Eight ARI and HumRRO staff members retranslated the anchors (high, moderate, and low performance) into dimensions, rated the levels of effectiveness, and provided written comments.

Based on that input, we revised the anchors. We also added an overall performance rating that uses a relative scale to the Army-wide PRS (see Figure 5.1).

We presented the revised Army-wide training and the MOS-specific training PRS to the Army Test Program Advisory Team (ATPAT)[13] for review. They made a few comments on the wording for the different scales, and we made edits based on their comments.

A. Overall Performance

Considering your evaluation of the Soldier on the dimensions important to successful performance, please rate the overall effectiveness of each Soldier compared to his/her peers.

1	2	3	4	5
Among the Weakest	**Below Average**	**Average**	**Above Average**	**Among the Best**
(in the bottom 20% of Soldiers)	(in the bottom 40% of Soldiers)	(better than the bottom 40% of Soldiers, but not as good as the top 40%)	(in the top 40% of Solders)	(in the top 20% of Soldiers)

Figure 5.1. Relative overall performance rating scale.

Army Life Questionnaire (ALQ)

The ALQ was designed to measure Soldiers' self-reported attitudes and experiences in training. The original form of the ALQ was developed for a prior ARI project (Van Iddekinge, Putka, & Sager, 2005) and based on those findings, it was modified slightly for use in the TOPS IOT&E. It focuses on first-term Soldiers' attitudes and experiences in initial military training (IMT) and includes 13 scales that cover (a) Soldiers' commitment and retention-related attitudes, and (b) Soldiers' performance and adjustment. Each ALQ scale was scored differently depending on the nature of the attribute being measured. The Army Physical Fitness Test (APFT) is a write-in item, and Training Achievements, Training Failures, and Disciplinary Incidents are simply a sum of the 'YES' responses. The remaining scales (see Table 5.3) are scored with Likert-type scales by computing a mean of the constituent item scores.

[13] The ATPAT is a group of senior non-commissioned officers (NCOs) originally established to provide guidance and support to earlier ARI enlisted research projects, and is continuing in this role for the Army Class project and the TOPS IOT&E. ATPAT membership has evolved, but generally has representatives from each MOS targeted in the research, G-1, Training and Doctrine Command (TRADOC), FORSCOM, and each of the components. Member names are listed in the acknowledgements of this report.

Table 5.3. ALQ Scales

Scale Name	Description	Example Item	Likert Scale Anchors
Affective Commitment	Measures Soldiers' emotional attachments to the Army.	I feel like I am part of the Army 'family.'	1 (strongly disagree) to 5 (strongly agree)
Normative Commitment	Measures Soldiers' feelings of obligation toward staying in the Army until the end of their current term of service.	I would feel guilty if I left the Army before the end of my current term of service.	1 (strongly disagree) to 5 (strongly agree)
Career Intentions	Measures intentions to re-enlist and to make the Army a career.	How likely is it that you will make the Army a career?	1 (strongly disagree) to 5 (strongly agree); 1 (not at all confident) to 5 (extremely confident); 1 (extremely unlikely to 5 (extremely likely)
Reenlistment Intentions	Measures Soldiers' intention to reenlist in the Army.	How likely is it that you will leave the Army after completing your current term of service?	1 (strongly disagree) to 5 (strongly agree)
Attrition Cognition	Measures the degree to which Soldiers think about attriting before the end of their first term.	How likely is it that you will complete your current term of service?	1 (strongly disagree) to 5 (strongly agree); 1 (never) to 5 (very often)
Army Life Adjustment	Measures Soldiers' transition from civilian to Army life	Looking back, I was not prepared for the challenges of training in the Army.	1 (strongly disagree) to 5 (strongly agree)
Army Civilian Comparison	Measures Soldiers' impressions of how Army life compares to civilian life.	Indicate how you believe conditions in the Army compare to conditions in a civilian job with regards to pay.	1 (much better in the Army) to 5 (much better in civilian life)
MOS Fit	Measures Soldiers' perceived fit with their MOS.	My MOS provides the right amount of challenge for me.	1 (strongly disagree) to 5 (strongly agree)
Army Fit	Measures Soldiers' perceived fit with their MOS.	The Army is a good match for me.	1 (strongly disagree) to 5 (strongly agree)

Administrative Criteria

Attrition is a broad category that encompasses involuntary and voluntary separations for a variety of reasons (e.g., underage enlistment, conduct, family concerns, sexual orientation, drugs or alcohol, performance, physical standards or weight, mental disorder, or violations of the Uniformed Code of Military Justice). Soldiers who were classified as attrits for reasons outside of their or the Army's control (e.g., death or serious injury incurred while performing one's duties) were excluded in our analyses. The reason for separation was determined by the Separation Program Designator (SPD) associated with the Soldier.

Data on IMT school performance and completion were extracted from the Army Training Requirements and Resources System (ATRRS) and the Resident Integrated Training Management System (RITMS) databases (see Chapter 2). ATRRS course information was used to determine (a) whether a Soldier graduated or was discharged during IMT and (b) the number of times he or she restarted during IMT. RITMS data, for those MOS that are providing data, were used to determine Soldiers' Advanced Individual Training (AIT) course grades. Given that each course has different grading procedures, the AIT course grade analysis variable was created by standardizing the grades within course. Due to restricted variance in the One Station Unit Training (OSUT) grades (i.e., all of the grades were pass/fail), these courses were excluded from the course grade analysis variable.

Training Criterion Measure Scores and Associated Psychometric Properties

Here we provide a review of the psychometric properties of the training criteria. Basic descriptive statistics are available for the full schoolhouse sample (n = 7,932, of which 7,700 had useable data for at least one criterion measure) and by MOS in Appendix C along with the intercorrelations. In this chapter we review the psychometric characteristics of the criterion measures estimated using only data from the Accession sample (i.e., Education Tier 1, non-prior service) that was used in the criterion-related validity analyses reported in Chapter 6 (n = 361 for schoolhouse IMT criteria, 1,050 for administrative IMT criteria, and 2,806 for attrition). Note, however, that the means, standard deviations, and reliability estimates are generally similar to those for the full schoolhouse sample.

Job Knowledge Tests (JKTs)

A single, overall raw score was computed for each JKT by summing the total number of points Soldiers earned across the final set of items retained for each JKT. All of the multiple-choice items were worth one point. Depending on the format of the non-traditional items (e.g., multiple response), they were worth one or more points. JKT records were flagged as not useable if the Soldier omitted more than 10% of the assessment items, took fewer than 5 minutes to complete the entire assessment[14], or chose an implausible response to one of the careless responding items. To facilitate comparisons across MOS, we computed a percent correct score based on the maximum number of points that could be obtained on each MOS test. For the

[14] The 5-minute criterion was established during the first in-unit phase of the Army Class project, which employs highly similar assessments administered via the same platform. See Knapp, Owens, and Allen (2010) for details.

criterion-related validity analyses, we converted the total raw score to a standardized score (or *z*-score) by standardizing the scores *within* each MOS.

Table 5.4 shows the percent correct scores, as well as internal consistency reliability estimates for the six MOS-specific and the WTBD JKTs. The mean percent correct score across all six MOS-specific tests was 69.1% versus 62.1% found in Army Class (Knapp & Heffner, 2009). Internal consistency reliability estimates were acceptable for those MOS with a useable sample size. Table C.5 in Appendix C shows the correlations between the various MOS JKT scores with the WTBD JKT score. The effect sizes range from small to moderate with all but the correlation with the 19K JKT significant, which has the smallest sample size (see Table C.1). These results suggest that the MOS-specific JKTs and the WTBD JKT each cover some unique content.

Table 5.4. Descriptive Statistics and Reliability Estimates for Training Job Knowledge Tests (JKTs) in the Applicant Sample

	n	*M*	*SD*	*Min*	*Max*	*α*
MOS-Specific Job Knowledge Test (JKT)						
11B/11C/11X/18X	134	58.28	8.90	34.78	79.35	.75
19K	1	78.00	--	78.00	78.00	n/a
31B	34	72.96	9.79	48.54	91.26	.82
68W	53	73.24	10.63	38.04	88.04	.87
88M	44	69.54	10.82	47.22	86.11	.78
91B	8	62.63	--	31.96	76.29	n/a
WTBD Job Knowledge	342	65.85	12.80	25.81	90.32	.65

Note. n/a = Internal consistency/coefficient alpha could not be computed or were inappropriate to compute (due to low sample size) for the scales/measures. Means represents percent correct. α = coefficient alpha. Results are limited to the Accession sample (non-prior service, Education Tier 1, AFQT Category IV and above).

Performance Rating Scales (PRS)

A single overall score was created for each Army-wide (AW) performance dimension and a composite of the MOS-specific performance rating scales (PRS). Computing these scores involved (a) computing the average of multiple ratings provided by cadre (if more than one rated the target Soldier) and (b) computing the mean of the individual scales that constitute the elements of a particular dimension. Approximately 24% of Soldiers were rated by more than one cadre member. The second step was only completed for the MOS-specific PRS, because each of the individual Army-wide scales represented a unique dimension. Overall mean ratings were calculated for every Soldier.[15] PRS data were flagged as unusable if the cadre member omitted more than 10% of the assessment items or indicated that he or she "Cannot Rate" the individual on more than 50% of the dimensions.

Descriptive statistics and estimates of internal consistency reliability for the Army-wide PRS dimensions and MOS PRS composite scores are shown in Table 5.5. Mean ratings are all above average, a common finding in research involving performance ratings. While the sample sizes (i.e., number of raters and ratees in the target sample) made the interrater reliability

[15] There were five dimensions on the 88M and 91B rating scales, seven dimensions on the 19K and 68W rating scales, eight dimensions on the 11B and 31B rating scales, and nine dimensions on the Army-wide rating scales.

computations inappropriate for the Accession sample, we computed them for the full schoolhouse sample and reported the results in Appendix C in Table C.3. To summarize, the interrater reliability estimates range from .08 to .24 for the AW scales in the full sample although the strength of the estimates varies by MOS with 91Bs having very good interrater reliability estimates and 88M having very poor interrater reliability estimates. We attribute the low coefficients to a few interrelated issues. First, the number of ratees per rater was rather high. It averaged 15.5 for the full schoolhouse sample. Second, most raters had very little variance in their ratings, perhaps reflecting their lack of familiarity with individual Soldiers. Third, these data collections were not proctored, while previous studies (e.g., Knapp & Heffner, 2009; 2010) had administered rating scales such as these in a proctored setting. Finally, the number of raters per target was small ($k < 2$), which reduces the magnitude of k-rater interrater reliability coefficients, such as the one reported in Appendix C.

In Table C.6 from Appendix C, we see that the correlations among the MOS PRS and the AW PRS are moderate to large, with all of them reaching significance. These results suggest there is more content overlap between the MOS PRS and the AW PRS than between the MOS JKTs and WTBD JKT. The AW scale that correlates the strongest with the MOS PRS is, not surprisingly, the *MOS Proficiency* scale. Whereas the MOS PRS that correlates most strongly with the AW PRS is 91B. The 91B PRS correlates most strongly with *Support for Peers, Peer Leadership, Common/Warrior Tasks,* and *MOS Proficiency* scales.

Table 5.5. Descriptive Statistics and Reliability Estimates for Training Performance Rating Scales (PRS) in the Applicant Sample

	n	*M*	*SD*	*Min*	*Max*	*α*
Army-Wide Performance Rating Scales						
Effort	174	4.83	1.14	1.00	7.00	n/a
Physical Fitness & Bearing	175	4.83	1.16	1.00	7.00	n/a
Personal Discipline	176	4.98	1.19	1.00	7.00	n/a
Commitment &Adjustment	176	5.00	1.14	1.00	7.00	n/a
Support for Peers	175	5.07	0.99	2.50	7.00	n/a
Peer Leadership	165	4.74	1.29	1.00	7.00	n/a
Common Warrior Tasks Knowledge and Skill	170	4.79	1.06	1.00	7.00	n/a
MOS Qualification Knowledge and Skill	161	4.84	1.05	1.00	7.00	n/a
Overall Performance Scale	170	3.50	0.74	1.00	5.00	n/a
MOS-Specific Performance Rating Composite Scores						
Total (combined across MOS)	163	4.68	0.88	2.71	7.00	n/a
11B/11C/11X/18X	66	4.77	0.86	3.00	6.63	.95
19K	--	--	--	--	--	n/a
31B	12	4.95	0.71	4.25	6.13	n/a
68W	52	4.39	0.86	2.71	6.29	.95
88M	26	4.66	0.57	3.80	6.20	.90
91B	7	5.74	1.37	3.00	7.00	n/a

Note. n/a = Internal consistency/coefficient alpha could not be computed or were inappropriate to compute (due to low sample size) for the scales/measures. Job knowledge scores are percent correct. Soldiers in this sample are non-prior service, Education Tier 1, AFQT Category IV or above Soldiers. The possible PRS scores are between 1 and 7 (highest), except for the Overall Performance Scale, which ranges from 1 to 5. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above).

Army Life Questionnaire (ALQ)

ALQ subscale scores are computed in most cases by taking the mean of all responses associated with each scale. The training failures, training achievement, and disciplinary action scales are computed by summing the total number of "yes" responses. Similar to the JKTs, a Soldier's ALQ data were flagged as unusable if the Soldier omitted more than 10% of the assessment items, took fewer than 5 minutes to complete the entire assessment, or chose an implausible response to the careless responding item.

Table 5.6 shows descriptive statistics and internal consistency reliability estimates for the training ALQ scores. The reliability estimates were good, ranging from .80 to .92. Mean scores were generally similar across MOS (see Table C.4 in Appendix C). Table C.3 shows that the subscales are generally positively correlated, with *Army Fit* having the strongest relationship with the other scales.

Table 5.6. Descriptive Statistics and Reliability Estimates for the ALQ in the Applicant Sample

Measure/Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
Affective Commitment	361	3.81	0.70	1.00	5.00	.87
Normative Commitment	361	4.07	0.79	1.00	5.00	.82
Career Intentions	361	3.06	1.11	1.00	5.00	.92
Reenlistment Intentions	361	3.49	1.06	1.00	5.25	.89
Attrition Cognition	361	1.61	0.72	1.00	5.00	.83
Army Life Adjustment	361	3.99	0.67	1.89	5.00	.85
Army Civilian Comparison	361	3.81	0.80	0.00	5.00	.80
MOS Fit	361	3.74	0.85	1.11	5.00	.92
Army Fit	361	3.98	0.62	1.00	5.00	.87
Training Achievement	361	0.39	0.59	0.00	2.00	n/a
Training Failure	361	0.38	0.60	0.00	2.00	n/a
Disciplinary Incidents	176	0.22	0.56	0.00	3.00	n/a
Last APFT Score	357	246.40	34.36	66.00	300.00	n/a

Note. n/a = Internal consistency/coefficient alpha could not be computed or were inappropriate to compute for the scales/measures. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above).

Administrative Criterion Data

For the first variable, *Graduation from IMT*, Soldiers who were discharged from the Army during IMT or failed to fully complete their training were coded as 0 (failure). Soldiers who completed IMT and graduated from AIT/OSUT were coded as 1 (graduate). Soldiers who failed to complete their IMT for nonacademic reasons that were administrative in nature and outside the Soldier's control were coded as missing (e.g., returned to unit for mobilization, unit recall, awaiting school start). Soldiers who had not had an opportunity to fully complete their IMT at the time the data were extracted were similarly excluded from our analyses. The second variable, *Number of Restarts During IMT*, was created by counting the total number of times a Soldier restarted during IMT.

Table 5.7 shows descriptive statistics for the graduation and restart IMT variables. The attrition rate was 6.1% for those Soldiers for whom 3-month attrition data were available. Table

C.10 shows that 19K Soldiers had the highest attrition rate (7.0%) and 68W Soldiers had the lowest (2.5%). Overall, 17.6% of the Soldiers restarted at least once during IMT. It is important to note that the IMT data retrieved from administrative sources were not mature. For example, although there were nearly 54,000 Soldiers in the sample, we retrieved attrition data on fewer than 3,000 and restart data on fewer than 1,100.

Table 5.7. Descriptive Statistics for Administrative Criteria Based on the Applicant Sample

Administrative Criterion	N [b]	N_{Attrit}	$\%_{Attrit}$
Three-Month Attrition[a]	2,806	170	6.1
Initial Military Training (IMT) Criteria	N [c]	$N_{Restarted}$	$\%_{Restarted}$
Restarted at Least Once During IMT	1,050	185	17.6
Restarted at Least Once During IMT for Pejorative Reasons	1,029	164	15.9
Restarted at Least Once During IMT for Academic Reasons	993	128	12.9
AIT School Grades	N [d]	*M*	*SD*
Overall Average (Unstandardized)	867	89.86	12.53
Overall Average (Standardized within MOS)	660	0.02	0.97

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above).
[a] Attrition results reflect Regular Army Soldiers only.
[b] N = number of Soldiers with 3-month attrition data at the time data were extracted. N_{Attrit} = number of Soldiers who attrited through 3 months of service. $\%_{Attrit}$ = percentage of Soldiers who attrited through 3 months of service [(N_{Attrit} /N) x 100].
[c] N = number of Soldiers with valid IMT data at the time data were extracted. N_{Failed} = number of Soldiers who failed at least once during IMT. $\%_{Failed}$ = percentage of Soldiers who failed at least once during IMT [(N_{Failed} /N) x 100].
[d] N = number of Soldiers with AIT school grade data. Standardized school grades were not computed for MOS with insufficient sample size (n < 15).

Summary

Three types of measures were adapted from previous Army research to validate the TAPAS: (a) job knowledge tests (JKTs), (b) performance rating scales (PRS), and (c) the Army Life Questionnaire (ALQ). The JKTs are completed by Soldiers in eight target MOS and measure MOS-specific and WTBD declarative and procedural knowledge. The PRS are completed by cadre and measure MOS-specific competence and Army-wide constructs such as effort and leadership. Finally, the ALQ asks Soldiers to complete self-report verifiable performance items (e.g., their APFT scores) and attitudinal items (e.g., adjustment to Army life). The scoring procedures were instrument-specific. In general, the criterion measures exhibited acceptable and theoretically consistent psychometric properties. The exception to this was the Army-wide and MOS-specific PRS, which exhibited high variable interrater reliability coefficients in the schoolhouse sample (see Appendix C). Results concerning these scales should be interpreted with caution. Additional criterion data, such as attrition, training restarts, and AIT course grades were gathered from administrative records.

CHAPTER 6: INITIAL EVIDENCE FOR THE PREDICTIVE VALIDITY AND CLASSIFICATION POTENTIAL OF THE TAPAS

D. Matthew Trippe, Joseph P. Caramagno, Matthew T. Allen, and Michael J. Ingerick
(HumRRO)

This chapter presents the results of analyses examining the potential of the TAPAS to improve enlisted Soldier selection and classification. At the time of these analyses, we only had schoolhouse criterion data for a small percentage of the applicant sample (0.7%) and administrative data for 6.7% of the applicant sample (see Table 2.3). Accordingly, the analyses we conducted focus on the TAPAS' potential to enhance new Soldier selection and classification and not on estimating the actual gains from its operational use. The results reported in this chapter should be treated as highly preliminary until criterion information can be gathered on a more representative sample. Predictive validity analyses assessing the TAPAS' potential for selection purposes are presented first, followed by classification-oriented analyses.

Predictive Validity

Analyses

To examine the TAPAS' potential to enhance new Soldier selection, we examined its incremental validity over the AFQT in predicting early first-term outcomes important to the Army. Consistent with the Army's personnel goals, we selected performance and retention-related outcomes that provided representative coverage of the criterion space. The criterion space for first-term Soldier performance can be specified using three higher-order domains (Campbell, Hanson, & Oppler, 2001; Campbell, McHenry, & Wise, 1990; Strickland, 2005). They are (a) can-do performance, which includes technical and soldiering proficiency; (b) will-do performance, which includes physical, interpersonal, and effort-related criteria; and (c) separation status, which includes attitudes that predict first-term Soldier attrition. These criterion measures were selected based on sample size considerations, psychometric properties, and coverage of each higher-order domain.

Our approach to analyzing the TAPAS' incremental predictive validity was consistent with previous evaluations of the measure or similar experimental non-cognitive predictors (Ingerick, Diaz, & Putka, 2009; Knapp & Heffner, 2009; 2010). In brief, this approach involved testing a series of hierarchical regression models, regressing each criterion measure onto Soldiers' AFQT scores in the first step, followed by their TAPAS scale scores in the second step. The resulting increment in the multiple correlation (ΔR) when the TAPAS scale scores were added to the baseline regression models served as our index of incremental validity.

For the continuously scaled criteria, these models were estimated using Ordinary Least Squares (OLS) regression. Specifically, estimating each model involved the following steps:

1. Estimating the observed (uncorrected) multiple correlation (R) for a baseline model focused on AFQT by regressing Soldiers' criterion scores onto their AFQT scores (i.e., AFQT only).

2. Estimating *R* for an alternative model containing the TAPAS by regressing Soldiers' criterion scores onto their AFQT and relevant TAPAS scale or composite scores (i.e., AFQT + TAPAS).

3. Calculating the increment in *R* (ΔR) by subtracting the uncorrected Step 1 *R* (AFQT only) from the uncorrected Step 2 *R* (AFQT + TAPAS).

Alternatively, logistic regression was used for the dichotomous criteria (3-month attrition, IMT graduation without a restart). At each step in the model, we estimated point-biserial correlations (r_{pb}) in place of the traditional pseudo *R* estimates to index incremental validity because of conceptual and statistical issues associated with these estimates. The point-biserial correlations reflected the correlation between a Soldiers' predicted probability of engaging in a behavior based on the predictors in the regression model and their actual behavior (e.g., attriting). Estimating these correlations involved the following steps:

1. Estimating a two-step hierarchal logistic regression model to obtain Soldiers' predicted probabilities on the criterion. Like the OLS models, Soldiers' AFQT scores were entered as the baseline predictor in the first step followed by their scores on the relevant TAPAS scales or composites as predictors in the second step.

2. Computing point-biserial correlations between the Soldiers' predicted probability of engaging in a behavior and their actual behavior based on the predictors in the regression model at each step. The incremental validity was computed by subtracting the point-biserial from Step 1 (AFQT only) from the point-biserial (AFQT + TAPAS) obtained from Step 2 (Δr_{pb}).

To supplement these incremental validity analyses, we also examined the predictive validity of the TAPAS at the scale level using bivariate and semi-partial correlations (controlling for AFQT). The semi-partial correlation provides information about the extent of influence on some outcome that is unique to a given predictor when multiple predictors influence the outcome by removing the effects of one predictor (i.e., AFQT) on the other (i.e., individual TAPAS scale) but not on the criterion (Cohen, Cohen, West, & Aiken, 2003). See Appendix D for the full set of bivariate and semi-partial correlations between the TAPAS composite and scale scores and all of the criteria described in Chapter 5. No corrections for multivariate range restriction or shrinkage were made because of the preliminary nature of these analyses.

As described in Chapter 3, three versions of TAPAS were administered in the TOPS IOT&E: (a) a 13-dimension computer adaptive version (13D-CAT), (b) a 15-dimension static version (15D-Static), and (c) a 15-dimension computer adaptive version (15D-CAT). Based on the results of our equivalence analysis (Chapter 4), we combined scores across the three versions when running the predictive validity analyses. To minimize scaling differences across the three versions, TAPAS scale or composite scores were standardized within version based on the population of interest (i.e., Education Tier 1, non-prior service, AFQT Category IV or above; see Chapter 4 for more details).

Criterion-Related Validity Evidence

Complete incremental validity analysis results can be found in Appendix D, Table D.1, while a subset of key criteria are presented in Table 6.1. The TAPAS predicted significant incremental variance beyond the AFQT for two criteria—Training Achievement and Last APFT Score. However, sample sizes were limited for a number of these criteria, especially the will-do performance criteria (with the exception of the Last APFT Score, sample sizes ranged from 118-129), which reduces the power to detect significant effects. Smaller sample sizes might also make estimates of multiple *R* unstable and difficult to interpret.

Consistent with previous research (e.g., Ingerick et al., 2009; Knapp & Heffner, 2009; 2010), the AFQT was generally more predictive of can-do performance-related criteria (*R*s ranged from .02 to .43) than will-do performance and retention-related criteria (*R*s ranged from .00 to .16). The incremental validity gains associated with the TAPAS were generally small to modest. Of the criteria in our subset, the only statistically significant incremental validity estimate was the Soldiers' self-reported APFT score.

Table 6.1. Incremental Validity Estimates for the TAPAS Scales over the AFQT for Predicting Select Performance- and Retention-Related Criteria

Criterion	*n*	AFQT Only *R* (r_{pb})	AFQT + TAPAS *R* (r_{pb})	Δ*R* (Δr_{pb})
WTBD Job Knowledge Test (WTBD JKT)	255	**.43**	**.51**	.08
MOS-Specific JKT	203	**.31**	**.41**	.09
IMT Exam Grade	544	**.23**	**.27**	.04
# of Restarts in IMT (ALQ)	670	.02	.17	.15
Last APFT Score (ALQ)	269	.03	**.34**	**.31**
Disciplinary Action (ALQ)	129	.05	.30	.25
Adjustment to Army Life (ALQ)	272	**.16**	**.32**	.16
Affective Commitment (ALQ)	272	.00	.22	.21
3-Month Attrition[a]	2,443	(.01)	**(.09)**	(.08)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. AFQT Only = Correlation between the AFQT and the criterion of interest. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT and the selected predictor measure with the criterion of interest. Δ*R* = Increment in *R* over the AFQT from adding the selected predictor measure to the regression model ([AFQT + TAPAS] – AFQT Only). *Point-biserial correlation (r_{pb})* = Observed point-biserial correlation between Soldiers' predicted probability of attriting and their actual attrition behavior. Large, positive r_{pb} values mean that the TOPS composite or scale performed well in predicting actual attrition. Results are limited to non-prior service, Education Tier 1, AFQT Category IV and above applicants. Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a]Attrition results include Regular Army Soldiers only.

The pattern of Δ*R*'s reported here were very similar to those found in the EEEM research. Knapp and Heffner (2010) reported incremental validities for three of the criteria shown in Table 6.1, including MOS-Specific Job Knowledge Tests (JKT) (Δ*R* = .03), Last APFT Score (Δ*R* = .28), and Affective Commitment (Δ*R* = .19). It should be noted that sample sizes found in EEEM were roughly three times larger than those reported here. The relationship between the AFQT and these matched performance- and retention-related criteria was also comparable in TOPS versus EEEM (MOS-Specific JKT, *R* = .44, $p < .05$; Last APFT Score, *R* = .05, *ns*; Affective Commitment, *R* = .07, *ns*).

Table 6.2 displays the bivariate and semi-partial correlations between the scores on the individual TAPAS scales/composites and the selected criterion measures. Although 85% of the bivariate correlations were not statistically significant ($p < .05$), there were a number of notable exceptions that were consistent with a theoretical understanding of the TAPAS scales and previous research (Knapp & Heffner, 2010). Specifically, Physical Conditioning was positively correlated with self-reported APFT score and Adjustment to Army Life, and negatively correlated with attrition and number of restarts. Intellectual Efficiency was positively correlated with WTBD JKT, IMT Exam Grade, and Adjustment to Army Life. A number of other TAPAS scales, including Achievement, Adjustment, and Optimism, also significantly predicted Adjustment to Army Life. Optimism also significantly predicted 3-month attrition. There were also a few other statistically significant correlations such as Generosity being negatively correlated with WTBD and MOS-specific job knowledge and Sociability being negatively correlated with IMT Exam Grade.

Examination of the scale-level incremental validity coefficients in Table 6.2 shows that this general pattern of results remained largely the same after controlling for AFQT, suggesting the TAPAS' impact on the criteria of interest is largely independent of AFQT. The notable exception was for Intellectual Efficiency, whose correlations with can-do performance criteria (WTBD JKT, IMT Exam Grade) dropped to nearly zero after controlling for AFQT. This finding makes theoretical sense and is consistent with prior research where Intellectual Efficiency has emerged as the TAPAS scale most strongly correlated with AFQT (Knapp & Heffner, 2010). In summary, this pattern of results suggests that the relationships between the TAPAS scales and the criteria are generally independent of AFQT.

Finally, we computed correlations between the TAPAS composite scores and the selected criteria by AFQT category to explore the potential influence these factors might have on our results (see Table 6.3). There were few statistically significant results, and sample sizes varied substantially across the AFQT categories. In some cases, sample sizes were as small as 64 cases,[16] suggesting potential instability in many of these estimates. Consistent with the scale-level results, the TAPAS can-do and will-do composites predicted Adjustment to Army Life and self-reported APFT scores at the highest rate. The can-do composite also predicted 3-month attrition at a significant rate for AFQT Category IIIA Soldiers. In general, the prediction rates tended to be highest for AFQT Category IIIA Soldiers.

[16] For the complete set of criteria, sample sizes dropped even further. The smallest sample sizes were generally associated with criteria assessed via performance rating scales (PRS) such as MOS-Specific performance and MOS Proficiency. Coefficients for AFQT Category IV Soldiers alone were not computed due to low sample size. See Appendix D.

Table 6.2. Bivariate and Semi-Partial Correlations between the TAPAS Scales and Selected Criteria

	Criteria									
	Can-do Performance				Will-do Performance		Retention			
TAPAS Dimensions	WTBD JKT *n* = 342	MOS-Specific JKT *n* = 274	IMT Exam Grade *n* = 660	# of Restarts (ALQ) *n* = 1,050	Disciplinary Incidents (ALQ) *n* = 176	Last APFT Score (ALQ) *n* = 357	Adjustment to Army Life (ALQ) *n* = 361	Affective Commitment (ALQ) *n* = 361	3-Month Attrition[b] *n* = 2,810	
Achievement	.04	(.00)	-.04	(-.06)	.01	(-.02)	-.01	(-.01)	**-.21**	**(-.20)**	.05	(.05)	**.13**	**(.12)**	.10	(.10)	.01	(.01)
Adjustment[a]	.12	(.07)	.08	(.04)	.01	(-.01)	**.09**	**(.09)**	-.08	(-.08)	.03	(.03)	**.18**	**(.17)**	-.04	(-.04)	.01	(.01)
Attention Seeking	-.04	(-.08)	-.05	(-.08)	.00	(-.02)	.00	(.00)	.05	(.06)	.00	(.00)	.00	(-.01)	.02	(.02)	.00	(.00)
Cooperation	-.06	(-.06)	.02	(.02)	.03	(.04)	.03	(.03)	.03	(.03)	.06	(.06)	-.02	(-.02)	.01	(.01)	-.01	(-.01)
Dominance	.02	(-.02)	-.11	**(-.14)**	.02	(.00)	-.04	(-.04)	-.03	(-.03)	.08	(.08)	.10	(.08)	.00	(.00)	-.02	(-.02)
Even Tempered	-.11	**(-.14)**	-.04	(-.06)	.01	(-.01)	.00	(.00)	-.04	(-.04)	-.10	(-.10)	.09	(.08)	.09	(.09)	-.01	(-.01)
Generosity	**-.17**	**(-.14)**	**-.18**	**(-.16)**	.01	(.03)	-.03	(-.03)	-.04	(-.04)	.06	(.06)	-.07	(-.06)	.07	(.07)	.01	(.01)
Intellectual Efficiency	**.20**	(.01)	.11	(-.03)	**.11**	(.01)	.05	(.04)	.00	(.03)	.00	(-.01)	**.18**	**(.12)**	-.01	(-.01)	-.01	(.00)
Non-delinquency	-.08	(-.08)	-.09	(-.08)	-.03	(-.03)	.00	(.00)	-.09	(-.09)	.08	(.08)	.02	(.02)	.04	(.04)	.00	(.00)
Optimism	.03	(.03)	-.03	(-.03)	.01	(.01)	-.02	(-.02)	-.06	(-.06)	.03	(.03)	**.12**	**(.11)**	.02	(.02)	**-.05**	**(-.05)**
Order	**-.13**	(-.06)	-.08	(-.02)	.02	(.07)	-.07	**(-.07)**	-.01	(-.02)	.03	(.03)	.02	(.05)	.02	(.02)	-.02	(-.02)
Physical Conditioning	.03	(.01)	.00	(-.01)	-.04	(-.05)	**-.07**	**(-.07)**	-.10	(-.10)	**.27**	**(.27)**	**.13**	**(.12)**	.00	(.00)	**-.04**	(-.04)
Self-Control[a]	-.12	**(-.11)**	-.07	(-.06)	.04	(.05)	.06	(.06)	.07	(.07)	.07	(.07)	-.03	(-.03)	**.14**	**(.14)**	-.01	(-.01)
Sociability	-.03	(.00)	-.07	(-.05)	**-.09**	**(-.08)**	-.03	(-.02)	.04	(.03)	.07	(.07)	.05	(.06)	.04	(.04)	-.03	(-.03)
Tolerance	-.09	(-.08)	-.09	(-.08)	.01	(.02)	-.04	(-.04)	-.05	(-.05)	.09	(.09)	.03	(.04)	**.11**	**(.11)**	.00	(-.01)
TAPAS Composites																		
Can-do Composite	.03	(-.07)	-.03	(-.10)	.04	(-.01)	.01	(.00)	-.14	(-.14)	.02	(.02)	**.19**	**(.16)**	.08	(.08)	-.02	(-.02)
Will-do Composite	-.04	(-.05)	-.04	(-.06)	-.03	(-.03)	-.03	(-.03)	**-.19**	**(-.19)**	**.12**	(.12)	**.14**	**(.14)**	.08	(.08)	-.02	(-.01)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Estimates in parentheses are semi-partial correlations between the TAPAS scales and the criterion of interest, controlling for AFQT. Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a] Adjustment and Self Control were included in the TAPAS 15-dimension versions (i.e., static and CAT) only. Sample sizes for these scales are smaller, ranging from 113 - 2,443.

[b] Attrition results include Regular Army Soldiers only.

Table 6.3. Correlations between TAPAS Composite Scores and Select Performance and Retention-Related Criteria

	AFQT Category							
	I-II		IIIA		IIIB		I-IV	
TAPAS Composite/Criterion	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*
TAPAS Can-Do Composite								
WTBD JKT	.00	162	.18	78	-.20	80	.03	342
MOS-Specific JKT	-.06	134	-.23	60	-.20	64	-.03	274
IMT Exam Grade	.01	310	-.11	150	.14	176	.04	660
# of Restarts (ALQ)	.01	454	.02	267	-.00	287	.01	1,050
Adjustment to Army Life (ALQ)	.08	168	**.25**	86	.12	85	**.19**	361
Last APFT Score (ALQ)	-.03	166	**.22**	85	-.00	85	.02	357
Affective Commitment (ALQ)	-.03	168	.19	86	.18	85	.08	361
3-Month Attrition[a]	.00	1,334	**-.10**	648	-.00	733	-.02	2,810
TAPAS Will-Do Composite								
WTBD JKT	-.01	162	.14	78	-.20	80	-.04	342
MOS-Specific JKT	-.13	134	-.05	60	.02	64	-.04	274
IMT Exam Grade	-.04	310	-.07	150	.04	176	-.03	660
# of Restarts (ALQ)	-.01	454	-.05	267	-.10	287	-.03	1,050
Adjustment to Army Life (ALQ)	.11	168	.07	86	.21	85	**.14**	361
Last APFT Score (ALQ)	.11	166	**.26**	85	.09	85	**.12**	357
Affective Commitment (ALQ)	.02	168	.05	86	.21	85	.08	361
3-Month Attrition[a]	.02	1,334	-.07	648	-.00	733	-.02	2,810

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Estimates in bold were statistically significant, $p < .05$ (two-tailed).
[a]Attrition results include Regular Army Soldiers only.

Classification Potential

Analyses

Because of the importance of maximizing person-job fit, the Army is interested in evaluating the TAPAS' potential for improving new Soldier classification in addition to examining its potential for entry-level selection (Ingerick, et al., 2009; Knapp, Owens, & Allen, 2010). We have typically analyzed a measure's classification potential using (a) Horst's (1954, 1955) index of differential validity (H_d) and (b) mean predicted criterion score (MPCS; De Corte, 2000)—two standard metrics for evaluating a measure's classification potential. However, we elected to examine the TAPAS' classification potential using a simpler set of metrics because of the preliminary nature of the present analyses and the limited amount of criterion data available at this stage. Future iterations of the TOPS IOT&E analyses will employ Horst's *d* and MPCS once sufficient criterion data are available. Accordingly, the results of the current analyses should be interpreted as preliminary.

In place of Horst's *d* and MPCS, we examined cross-MOS differences in TAPAS score profiles and predictive validity estimates. Like H_d and MPCS, these alternative metrics summarize cross-job variability or differences in predictor-criterion scores. All other factors being equal, classification potential will be low if there is little cross-MOS variability (or differences) in scores on the predictor measure or in their relations to selected criteria. This is because the lack of cross-MOS differences means that it makes no practical difference where new Soldiers are classified; Soldiers would be expected to perform equally well or to persist equally as long in any given MOS.

Cross-MOS Differences in TAPAS Score Profiles

Cross-MOS differences in TAPAS score profiles were examined by computing the overall average root mean squared difference (RMSD) in TAPAS scale scores across MOS. The average RMSD was computed across all TAPAS scales as

$$RMSD = \sqrt{\frac{\sum_{d=1}^{D} \sum_{j,k=1,\, j \neq k}^{J} (\overline{x}_j - \overline{x}_k)_d^2}{n_D \, n_{J-1}}}$$

where *d* represents a TAPAS dimension, *j* represents an MOS and *k* represents an MOS different from *j*. In addition to computing the overall average RMSD across all TAPAS scales, we also calculated the RMSDs for each TAPAS scale, as well as for the two TAPAS composites. RMSD values computed by TAPAS scale (or composite) were calculated as

$$RMSD = \sqrt{\frac{\sum_{j,k=1,\, j \neq k}^{J} (\overline{x}_j - \overline{x}_k)^2}{n_{J-1}}} .$$

Conceptually, this metric provides an index of how much the mean TAPAS scale scores differ, on average, among the MOS being sampled. The larger the average RMSD value, the greater the differences, on average, in mean TAPAS scores across the MOS sampled.[17] Bigger cross-MOS differences in TAPAS score profiles mean that Soldiers with different score profiles are more likely to be attracted (or to gravitate) to select MOS than others. Although focused on the predictor-side, these differences provide evidence for a measure's classification potential.

Tables 6.4 and 6.5 summarize the average RMSDs for the target and an expanded sample of MOS, respectively. These additional MOS were selected because they (a) had relatively high volumes of TAPAS data and (b) represented career fields or had aptitude requirements different from those covered by the eight target MOS. The additional MOS selected were 21B (Combat Engineer), 35F (Intelligence Analyst), and 92G (Food Service Specialist).

Table 6.4 indicates that there were cross-MOS differences in mean TAPAS score profiles across the eight target MOS. However, these differences were generally small in magnitude. RMSD values ranged from .10 (88M) to .16 (68W) when computed across all TAPAS scales. These values can be placed in perspective by comparing them to the RMSD values computed for the ASVAB subtests found at the bottom of Table 6.4. ASVAB subtests were chosen over the Aptitude Area (AA) composites for the comparative index because (a) AA composite scores are empirically keyed to criteria that are not necessarily the same criteria the TAPAS was designed to predict; (b) ASVAB subtests and TAPAS scales exist at similar levels in the construct space; and (c) TAPAS scales are more amenable to comparisons to ASVAB subtests than they are to AA composites. RMSD values for the TAPAS scales were appreciably smaller than those observed in the ASVAB. This is not entirely unexpected given that ASVAB scores play a

[17] The average RMSD can only attain positive values because the mean score differences are squared.

significant role in classifying new Soldiers into MOS. Nevertheless, cross-MOS differences in mean TAPAS score profiles were comparatively small in magnitude, which suggests the potential for gains in classification efficiency may also be small relative to the ASVAB.

Table 6.4. Average Root Mean Squared Differences in Mean TAPAS Scale Score Profiles for the Eight Target MOS

Composite/ Scale Score Profile	11B	19K	25U	31B	42A	68W	88M	91B	Avg	Min	Max
All TAPAS Scales	.14	.14	.12	.12	.15	.16	.10	.13	.13	.10	.16
TAPAS Scale											
Achievement	.10	.08	.10	.08	.08	.13	.07	.11	.09	.07	.13
Adjustment	.18	.25	.13	.13	.27	.13	.13	.12	.17	.12	.27
Attention Seeking	.13	.10	.13	.11	.11	.18	.10	.18	.13	.10	.18
Cooperation	.10	.10	.17	.10	.08	.09	.08	.13	.11	.08	.17
Dominance	.15	.11	.14	.20	.12	.14	.12	.20	.15	.11	.20
Even Tempered	.07	.13	.07	.12	.08	.13	.08	.08	.10	.07	.13
Generosity	.18	.17	.12	.12	.23	.21	.12	.13	.16	.12	.23
Intellectual Efficiency	.17	.21	.16	.15	.19	.38	.17	.20	.20	.15	.38
Non-Delinquency	.08	.06	.10	.10	.06	.06	.08	.12	.08	.06	.12
Optimism	.07	.06	.11	.06	.08	.05	.05	.05	.07	.05	.11
Order	.11	.13	.08	.10	.15	.10	.11	.12	.11	.08	.15
Physical Conditioning	.28	.18	.13	.21	.20	.13	.14	.14	.18	.13	.28
Self-Control	.05	.12	.06	.05	.06	.08	.06	.09	.07	.05	.12
Sociability	.05	.06	.07	.07	.04	.06	.04	.07	.06	.04	.07
Tolerance	.15	.12	.12	.14	.23	.21	.11	.16	.16	.11	.23
TAPAS Composite											
Can-Do Composite	.10	.10	.09	.09	.11	.23	.11	.15	.12	.09	.23
Will-Do Composite	.09	.06	.05	.07	.09	.06	.06	.06	.07	.05	.09
All ASVAB Subtests	.32	.34	.31	.28	.55	.54	.32	.35	.38	.28	.55
ASVAB Subtests											
Arithmetic Reasoning	.27	.28	.31	.28	.39	.67	.32	.34	.36	.27	.67
Auto & Shop Information	.41	.49	.40	.35	.79	.35	.33	.44	.45	.33	.79
Electronics Information	.36	.38	.29	.31	.73	.46	.31	.30	.39	.29	.73
General Science	.33	.35	.29	.28	.59	.58	.34	.32	.39	.28	.59
Mechanical Comprehension	.36	.36	.28	.29	.69	.47	.30	.29	.38	.28	.69
Math Knowledge	.23	.32	.36	.23	.24	.50	.27	.31	.31	.23	.50
Paragraph Comprehension	.27	.26	.28	.26	.35	.63	.34	.38	.35	.26	.63
Word Knowledge	.27	.26	.28	.26	.38	.60	.30	.40	.34	.26	.60

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis. TAPAS sample sizes by MOS are: 11B = 2,107, 19K = 158, 25U= 290, 31B = 907,42A = 410, 68W = 1,139, 88M = 1,149, 91B = 775. ASVAB sample sizes by MOS are 11B = 1,746, 19K = 151, 25U = 231, 31B = 680, 42A = 345, 68W = 993, 88M = 1,036, 91B = 654. The last three columns represent the Average, Minimum and Maximum RMSD values presented in the table.

Examining RMSD values by TAPAS scales reveals that the magnitude of these cross-MOS differences varied by scale. For example, scores on the Adjustment, Dominance, Intellectual Efficiency, and Physical Conditioning scales exhibited larger cross-MOS differences, on average, than did scores on the Self-Control, Sociability, and Optimism scales. Examination of Table D.7 in Appendix D provides further insight into the source and direction of these differences. The larger RMSD values observed for the Physical Conditioning scale appear to be driven by 11B and 31B, whose mean scores were higher than other MOS. This is consistent with the occupational requirements of these MOS, which tend to be among the more physically demanding of the target

MOS. We expected 19K to exhibit relatively higher Physical Conditioning scores along with 11B and 31B, but this was not observed in the present sample. This finding could be attributable to the fact that the 19K sample available, which is the smallest included in the analyses, was too small to exhibit the expected profile. The larger RMSD values observed for Intellectual Efficiency scale appears driven by the relatively higher scores observed for 68W. This is consistent with the finding that 68W Soldiers have higher ASVAB scores relative to other MOS in the target samples. Cross-MOS differences in the Dominance scale appear to be driven by relatively high scores in 31B and relatively low scores in 91B. High Dominance is conceptually consistent with the occupational profile of 31B. Adjustment scale differences result from relatively high scores observed in the 11B and 19K samples and relatively low scores observed in the 42A sample. Adjustment may be of greater importance in combat MOS than administrative occupations because of combat related stressors. With regards to the TAPAS composites, scores on the TAPAS can-do composite yielded higher RMSD values, on average, than those on the will-do composite. In summary, these findings suggest that the TAPAS has classification potential. Pursuing the more sophisticated H_d and MPCS analyses in the future will provide a more definitive evaluation and estimate of its potential.

Table 6.5 reports the RMSD values based on the expanded sample of MOS, across all TAPAS scales and by TAPAS scale (or composite). RMSD values computed on the ASVAB are again presented to provide a reference or baseline against which the TAPAS results can be meaningfully compared. Overall, the addition of MOS resulted in somewhat larger cross-MOS differences in mean TAPAS score profiles than those observed for the eight target MOS. RMSD values ranged from .12 (88M) to .19 (35F, 92G) when computed across all TAPAS scales. RMSD values for the TAPAS were again relatively smaller than those observed for the ASVAB. Consistent with the previous analyses, scores on the Intellectual Efficiency and Physical Conditioning scales demonstrated higher cross-MOS differences, on average, than scores from the other TAPAS scales. Scores on the Generosity scale also exhibited relatively higher levels of cross-MOS differences than the other scales. Cross-MOS score differences in Intellectual Efficiency appear to be driven by 68W and 35F, who scored higher than other MOS on this scale. This is consistent with the observation that these MOS also have relatively higher ASVAB scores than other MOS in the sample. Physical Conditioning differences continue to be driven by 11B and 31B, arguably among the more physically demanding of the MOS sampled. Cross-MOS score differences on the Generosity scale appears to be driven by 42A, 68W, and 92G—the more service-oriented of the MOS sampled. Scores on the Sociability and Optimism scales continued to evidence the lowest cross-MOS differences, on average. As with the eight targeted MOS, scores on the TAPAS can-do composite exhibited larger mean differences, on average, than scores on the will-do composite.

Cross-MOS Differences in Predictive Validity Estimates

To further evaluate the TAPAS' classification potential, we also examined cross-MOS differences in predictive validity estimates in addition to differences in TAPAS score profiles. The results of these analyses were intended to complement those from the TAPAS score profile analyses. Whereas the preceding analyses focused on scores on the predictor side, the current analyses incorporate scores on relevant criteria. In doing so, these analyses provide a more direct assessment of the TAPAS' potential to differentially predict how well Soldiers will perform or persist among a targeted sample of MOS—all other factors being equal, the greater the differential prediction, the higher the TAPAS' classification potential.

Table 6.5. Average Root Mean Squared Differences in Mean TAPAS Scale Score Profiles for the Expanded Sample of MOS

Composite/ Scale Score Profile	11B	19K	25U	31B	42A	68W	88M	91B	21B	35F	92G	Avg	Min	Max
All TAPAS Scales	.15	.14	.12	.14	.16	.16	.12	.14	.15	.19	.19	.15	.12	.19
TAPAS Scale														
Achievement	.12	.09	.10	.09	.09	.14	.08	.11	.08	.14	.18	.11	.08	.18
Adjustment	.18	.25	.12	.12	.25	.12	.12	.12	.14	.12	.19	.16	.12	.25
Attention Seeking	.14	.11	.13	.12	.11	.19	.11	.17	.12	.12	.22	.14	.11	.22
Cooperation	.10	.08	.16	.10	.07	.08	.08	.13	.07	.07	.11	.10	.07	.16
Dominance	.17	.12	.14	.21	.13	.15	.13	.20	.13	.22	.22	.17	.12	.22
Even Tempered	.09	.11	.08	.14	.09	.11	.10	.09	.09	.15	.07	.10	.07	.15
Generosity	.19	.18	.14	.14	.25	.23	.14	.15	.26	.18	.27	.19	.14	.27
Intellectual Efficiency	.20	.26	.19	.20	.24	.36	.22	.25	.19	.48	.25	.26	.19	.48
Non-Delinquency	.11	.09	.10	.11	.09	.09	.11	.15	.15	.19	.13	.12	.09	.19
Optimism	.09	.07	.11	.08	.10	.07	.07	.07	.09	.07	.17	.09	.07	.17
Order	.12	.13	.09	.11	.16	.11	.12	.13	.16	.10	.19	.13	.09	.19
Physical Conditioning	.31	.16	.13	.23	.18	.13	.14	.13	.15	.16	.24	.18	.13	.31
Self-Control	.07	.11	.08	.07	.07	.11	.07	.11	.10	.15	.10	.09	.07	.15
Sociability	.09	.07	.08	.10	.07	.09	.08	.08	.11	.19	.07	.09	.07	.19
Tolerance	.17	.14	.13	.15	.22	.21	.13	.17	.23	.15	.23	.18	.13	.23
TAPAS Composites														
Can-Do Composite	.13	.14	.13	.13	.15	.23	.15	.19	.13	.33	.19	.17	.13	.33
Will-Do Composite	.09	.06	.06	.07	.09	.06	.07	.07	.07	.11	.06	.07	.06	.11
All ASVAB Subtests	.34	.37	.33	.31	.57	.53	.35	.38	.35	.56	.50	.42	.31	.57
ASVAB Subtests														
Arithmetic Reasoning	.32	.34	.35	.31	.44	.66	.37	.39	.32	.71	.52	.43	.31	.71
Auto Shop	.43	.50	.40	.33	.77	.37	.34	.46	.48	.35	.65	.46	.33	.77
Electronics Information	.36	.38	.31	.35	.73	.46	.32	.31	.38	.44	.57	.42	.31	.73
General Science	.32	.34	.30	.32	.61	.55	.37	.34	.32	.55	.45	.41	.30	.61
Mechanical Comprehension	.37	.37	.31	.30	.70	.47	.33	.31	.36	.50	.58	.42	.30	.70
Math Knowledge	.28	.37	.36	.32	.30	.49	.33	.36	.28	.65	.37	.37	.28	.65
Paragraph Comprehension	.29	.29	.30	.27	.39	.61	.38	.42	.29	.63	.42	.39	.27	.63
Word Knowledge	.28	.28	.29	.30	.41	.57	.33	.44	.28	.60	.36	.38	.28	.60

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis. TAPAS sample sizes by MOS are: 11B = 2,107, 19K = 158, 25U = 290, 31B = 907, 42A = 410, 68W=1139, 88M = 1,149, 91B = 775, 21B = 572, 35F =338, 92G = 487. ASVAB sample sizes by MOS are: 11B = 1,746, 19K = 151, 25U = 231, 31B = 680, 42A = 345, 68W = 993, 88M = 1,036, 91B = 654, 21B = 498, 35F =314, 92G = 440. The last three columns represent the Average, Minimum and Maximum RMSD values presented in the table.

Similar to the preceding analyses, cross-MOS differences in predictive validity estimates were measured by computing an average RMSD in these estimates among the MOS sampled. The predictive validity estimates that served as input to this metric were based on seven selected criterion measures: (a) 3-month attrition, (b) graduation from AIT, (c) MOS specific JKT scores (standardized within MOS), (d) Warrior Tasks and Battle Drills (WTBD) JKT scores, (e) cadre ratings of MOS specific performance, (f) perceived MOS fit, and (g) attrition cognitions. These criterion measures were selected based on sample size considerations and have been used in prior classification analyses of the TAPAS and similar experimental predictor measures (Ingerick et al., 2009; Knapp et al., 2010). The average RMSD was calculated as

$$RMSD = \sqrt{\frac{\sum_{d=1}^{D} \sum_{j,k=1,\, j \neq k}^{J} (r_j - r_k)_d^2}{n_D \, n_{J-1}}}$$

where *d* represents a TAPAS dimension, *j* represents an MOS and *k* represents an MOS different from *j*. Computationally, this formula is similar to the RMSD formula used in the preceding mean score profile analyses, except in this case the primary inputs to the formula are predictive validity estimates (*r*'s) and not mean scores. Conceptually, this metric provides an index of how much the predictive validity estimates differ, on average, among the MOS being sampled. Larger RMSD values reflect greater differences, on average, in predictive validity estimates across the MOS sampled.[18]

As in the preceding analyses, we also calculated RMSDs by TAPAS scale and the two TAPAS composites. RMSD values by TAPAS scale (or composite) were computed using a simplified version of the above formula:

$$RMSD = \sqrt{\frac{\sum_{j,k=1,\, j \neq k}^{J} (r_j - r_k)_d^2}{n_{J-1}}}$$

Table 6.6 summarizes the RMSDs in predictive validity estimates for five target MOS for the TAPAS as a whole and by scale (or composite). Overall, Table 6.6 indicates that there were cross-MOS differences in predictive validity estimates. However, the magnitude of those differences varied by MOS and scale. RMSD values based on the full set of TAPAS scales ranged from .26 (11B) to .38 (91B). RMSD values for the TAPAS are comparable in magnitude to those computed on the ASVAB, suggesting that the variability in predictive validity estimates across MOS is similar or even slightly greater in the TAPAS. With respect to the individual scales, scores from the Adjustment, Intellectual Efficiency, and Optimism scales tended to demonstrate the biggest cross-MOS differences, on average, while scores on the Cooperation, Generosity, and Tolerance scales generally exhibited the smallest differences. Recall that the Adjustment and Intellectual Efficiency scales also

[18] Schoolhouse criterion data were not yet available for 19K, 25U, and 42A Soldiers, so they are not included in this analysis.

demonstrated relatively large cross-MOS differences in the mean profile analysis of the target MOS[19]. It is logically consistent that the variability observed in the mean profile analysis allows for more potential variability in predictive validity estimates. The Physical Conditioning scale, which emerged in the previous analyses as exhibiting relatively large cross-MOS mean differences, ranks in the middle with respect to the average RMSD value for predictive validity estimates. Nevertheless, it may be the case that Physical Conditioning is less relevant to the criterion variables involved in the present analysis.

Table 6.6. Average Root Mean Squared Differences in Predictive Validity Estimates for Five Target MOS

Composite/ Scale Score Profile	11B	31B	68W	88M	91B	Avg	Min	Max
All TAPAS Scales	.26	.31	.29	.27	.38	.30	.26	.38
TAPAS Scale								
Achievement	.25	.29	.28	.24	.34	.28	.24	.34
Adjustment	.32	.36	.35	.33	.56	.38	.32	.56
Attention Seeking	.26	.30	.36	.27	.39	.32	.26	.39
Cooperation	.21	.25	.22	.23	.34	.25	.21	.34
Dominance	.26	.27	.23	.25	.42	.29	.23	.42
Even Tempered	.27	.38	.30	.28	.43	.33	.27	.43
Generosity	.19	.20	.19	.19	.26	.21	.19	.26
Intellectual Efficiency	.33	.40	.33	.31	.54	.38	.31	.54
Non-Delinquency	.25	.36	.27	.29	.39	.31	.25	.39
Optimism	.30	.33	.36	.31	.47	.35	.30	.47
Order	.21	.32	.22	.23	.30	.26	.21	.32
Physical Conditioning	.28	.28	.33	.27	.37	.31	.27	.37
Self-Control	.25	.32	.30	.32	.24	.29	.24	.32
Sociability	.27	.30	.29	.25	.26	.27	.25	.30
Tolerance	.18	.20	.22	.23	.24	.21	.18	.24
TAPAS Composites								
Can-Do Composite	.36	.35	.39	.33	.43	.37	.33	.43
Will-Do Composite	.29	.41	.34	.30	.40	.35	.29	.41
All ASVAB Subtests	.20	.21	.25	.26	.25	.23	.20	.26
ASVAB Subtests								
Arithmetic Reasoning	.20	.26	.26	.29	.33	.27	.20	.33
Auto Shop	.22	.25	.35	.25	.23	.26	.22	.35
Electronics Information	.21	.20	.27	.30	.23	.24	.20	.30
General Science	.17	.22	.23	.21	.22	.21	.17	.23
Mechanical Comprehension	.21	.22	.30	.28	.24	.25	.21	.30
Math Knowledge	.16	.18	.16	.23	.25	.20	.16	.25
Paragraph Comprehension	.17	.20	.19	.22	.28	.21	.17	.28
Word Knowledge	.22	.16	.21	.26	.21	.21	.16	.26

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis. Criterion variable sample size ranges by MOS are 11B = 61-673, 31B =12-63, 68W = 9-160, 88M = 27-103, 91B = 7-82. Cadre ratings of MOS specific performance generally account for the lower end of the *n* range. The last three columns represent the Average, Minimum and Maximum RMSD values presented in the table.

[19] Note that analysis of predictive validity estimates was conducted on a subset of the Soldiers analyzed in the analysis of mean scores.

Cross-MOS differences in predictive validity estimates were similar in size based on scores from the can-do and will-do composites. Specific factors underlying the cross-MOS differences in predictive validity estimates were difficult to determine given the highly aggregate nature of these analyses. Nevertheless, it may be that the Adjustment and Optimism scales are of varying importance depending on the rigors or stressors associated with the MOS under consideration. Those Soldiers in more physically and psychologically demanding MOS may be more resilient as a result of more adaptable and positive personality attributes. The intellectual efficiency scale may be serving as a proxy for cognitive aptitude, and thus its individual potential for incremental classification potential beyond the ASVAB could be limited. The reader is cautioned against drawing firm conclusions based on individual RMSD values because the sample sizes for some MOS-criterion measure combinations were low. Nevertheless, the overall pattern of results suggests that TAPAS scores evidence differential prediction (or validity) that could enhance new Soldier classification over the ASVAB.

Summary and Conclusion

In this chapter, we presented preliminary results regarding the TAPAS' potential to supplement existing enlisted Soldier selection and classification systems. This was accomplished by examining the results of the validity and classification-oriented analyses in relation to the Army's primary measure for accomplishing these tasks—the ASVAB.

The results of the selection-oriented analyses suggest that the individual TAPAS scales significantly predict a number of criteria of interest. In addition, many of these correlations were theoretically consistent with expectations. Most notably, the Physical Conditioning scale predicted Soldiers' self-reported APFT scores, number of restarts, adjustment to Army life, and 3-month attrition. The Optimism scale also significantly predicted 3-month attrition. Intellectual Efficiency predicted scores on the WTBD JKT and IMT Exam Grades. A number of scales (Achievement, Adjustment, Intellectual Efficiency, Physical Conditioning, and Optimism) predicted the Adjustment to Army Life scale. These results are consistent with both theoretical descriptions of these scales and previous research (Ingerick et al., 2009; Knapp & Heffner, 2010) supporting these scales' use in an operational setting.

With regard to classification potential, the results of the RMSD values on the mean differences for the overall TAPAS were comparatively smaller than those observed in the ASVAB. The magnitude of the differences varied by TAPAS scale, however, often in ways that are consistent with a theoretical understanding of the scale and the MOS. For example, the means for Physical Conditioning were higher for some of the more physically-oriented MOS, such as 11B and 31B. The mean for the Intellectual Efficiency scale was highest for 68W, the most cognitively-oriented MOS in the sample. The results of the RMSD on the predictive validity estimates found that the Adjustment, Intellectual Efficiency, and Optimism scales generally exhibited the largest differences across MOS.

Taken together, these results suggest that, while the magnitude of the validity and classification coefficients are not as large as those found in the experimental EEEM research (Knapp & Heffner, 2010), the TAPAS holds promise for both selection and classification-oriented purposes. Many of the scale-level coefficients are consistent with a theoretical

understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. However, given the restricted nature of the matched criterion sample, these results should be considered highly preliminary. This is particularly true for the PRS, which exhibited highly variable interrater reliabilities (see Appendix C) and had low sample sizes. Future analyses should expand on these results by examining operational applications of the TAPAS, such as developing new selection and classification composites and determining the effect of various cut scores.

CHAPTER 7: SUMMARY AND A LOOK AHEAD

Deirdre J. Knapp (HumRRO), Tonia S. Heffner and Leonard A. White (ARI)

Summary of the TOPS IOT&E Method

The Army is conducting an initial operational test and evaluation (IOT&E) of the Tier One Performance Screen (TOPS). The TOPS assessments, including the Tailored Adaptive Personality Assessment Screen (TAPAS), and soon the Work Preferences Assessment (WPA), are being administered to non-prior service applicants testing at MEPS locations.

To evaluate the TAPAS and WPA, the Army is collecting training criterion data on Soldiers in selected MOS as they complete their Initial Military Training (IMT). The criterion measures include job knowledge tests (JKTs); an attitudinal person-environment fit assessment, the Army Life Questionnaire (ALQ), and performance rating scales (PRS) completed by the Soldiers' cadre members. Course grades and completion rates are obtained from administrative records for all Soldiers, regardless of MOS.

Two waves of in-unit job performance data collection are also planned at approximately 18 month intervals, both of which will attempt to capture Soldiers from across all MOS who completed the TAPAS (and WPA) at entry. These measures will again include JKTs, the ALQ, and supervisor ratings. Finally, the separation status of all Soldiers who took the TAPAS at entry is being tracked throughout the course of the research.

The plan is to construct analysis datasets and conduct validation analyses at 6-month intervals throughout the three-year IOT&E period. In addition to updating extant criterion measures for the planned two waves of in-unit criterion data collection, we will develop MOS-specific measures (both training and in-unit) for two occupations – Signal Support Specialist (25U) and Human Resources Specialist (42A).

Summary of Initial Evaluation Results

A staggered schedule for getting schoolhouse testing underway along with the fact that there is generally an appreciable delay between when individuals take pre-enlistment tests and when they access into the Army resulted in small samples on which to conduct validation analyses. Thus, the selection and classification-oriented analyses reported here must be viewed with considerable caution.

TAPAS Construct Validity

The three versions of the TAPAS (13D-CAT, 15D-Static, and 15D-CAT) were consistent with one another in terms of their means, standard deviations, and patterns of intercorrelations. The two computer-adaptive versions of the TAPAS were particularly similar. Some of the TAPAS scales appeared more similar across the research and operational settings than others. The patterns of relations between TAPAS scales and individual difference variables (AFQT

scores, race, ethnicity, and gender), however, were generally consistent from the EEEM to TOPS settings. Keeping in mind that previous research has shown large differences between the experimental and operational use of temperament measures (White et al., 2008), these results suggest that the use of the TAPAS in an operational setting is promising.

Validity for Soldier Selection

The results of the selection-oriented analyses suggest that the individual TAPAS scales significantly predict a number of criteria of interest. In addition, many of these correlations were theoretically consistent with expectations. Most notably, the Physical Conditioning scale predicted Soldiers' self-reported APFT scores, number of restarts, adjustment to Army life, and 3-month attrition. The Optimism scale also significantly predicted 3-month attrition. Intellectual Efficiency predicted scores on the Warrior Tasks and Battle Drills (WTBD) JKT and initial military training (IMT) Exam Grades. A number of scales (Achievement, Adjustment, Intellectual Efficiency, Physical Conditioning, and Optimism) predicted the Adjustment to Army Life scale. These results are consistent with both theoretical descriptions of these scales and previous research (Ingerick et al., 2009; Knapp & Heffner, 2010) supporting these scales' use in an operational setting. Given that some of the scales are not included in either the can-do or will-do composites (e.g., Adjustment), but did predict aspects of Soldier performance, future work will develop more comprehensive selection-oriented composites.

Potential for Soldier Classification

With regard to classification potential, the results of the RMSD values on the mean differences for the overall TAPAS were comparatively smaller than those observed in the ASVAB. The magnitude of the differences varied by TAPAS scale, however, often in ways that are consistent with a theoretical understanding of the scale and the MOS. For example, the means for Physical Conditioning were higher for more physically-oriented MOS, such as 11B and 31B. The mean for the Intellectual Efficiency scale was highest for 68W, the most cognitively-oriented MOS in the sample. The results of the RMSD on the predictive validity estimates found that the Adjustment, Intellectual Efficiency, and Optimism scales generally exhibited the largest differences across MOS.

Taken together, these early evaluation results suggest that, while the magnitude of the validity and classification coefficients are not as large as those found in the experimental EEEM research (Knapp & Heffner, 2010), the TAPAS holds promise for both selection and classification-oriented purposes. Many of the scale-level coefficients are consistent with a theoretical understanding of the TAPAS scales, suggesting that the scales are measuring the characteristics that they are intended to measure. However, given the restricted nature of the matched criterion sample, these results should be considered highly preliminary. Future analyses should expand on these results by examining operational applications of TAPAS, such as developing new selection and classification composites and determining the effect of various cut scores.

A Look Ahead

The second set of TOPS evaluation analyses will be conducted early in CY2011 based on data collected through December 2010. The sample sizes for this next evaluation are expected to be considerably larger, thus supporting additional analyses (e.g., re-examination of how the will-do and can-do TAPAS composite scores are constructed) and yielding more generalizable results. At that point, data analyses will still be restricted to IMT and separation criteria and exclude MOS-specific schoolhouse criteria for two target MOS: 25U and 42A. Subsequent iterations of the evaluation analyses will introduce MOS-specific criterion data for these two MOS and in-unit performance data as they become available.

The analyses in this report were restricted to Education Tier 1 Soldiers (high school degree graduates) because (a) they are the focus of the original TOPS concept and (b) this allows relatively direct comparison of these results to those obtained in a more purely research setting (i.e., the *Expanded Enlistment Eligibility Metrics* project). Because the Army may wish to consider alternative selection models, just as it is likely to want the composite scores re-examined, future evaluations might include Soldiers in other Education Tiers.

Readers should thus look forward to a series of five more reports, published at approximately 6-month intervals, that document the method and findings of the Army TOPS IOT&E.

REFERENCES

Allen, M.T., Cheng, Y.A., Putka, D.J., Hunter, A., & White L. (2010). Analysis and findings. In D.J. Knapp & T.S. Heffner (Eds.). *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43,* 313-333.

Campbell, J., Hanson, M. A., & Oppler S. H. (2001). Modeling performance in a population of jobs. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification.* Hillsdale, NJ: Erlbaum.

Campbell, J.P., & Knapp, D.J. (Eds.) (2001). *Exploring the limits in personnel selection and classification.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chernyshenko, O.S., Stark, S., Drasgow, F., & Roberts, B.W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19,* 88-106.

Chernyshenko, O.S., Stark, S., Woo, S., & Conz, G. (2008, April). *Openness to Experience: Its facet structure, measurement and validity*. Paper presented at at the 23nd annual conference for the Society of Industrial and Organizational Psychologists. New Orleans, LA.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd edition*. Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., Aiken, L. S., & West, S. G., (2003). *Applied multiple regression/correlation analysis for the behavioral and social sciences* (3rd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.

Collins, M., Le, H., & Schantz, L. (2005). Job knowledge criterion tests. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (Technical Report 1168) (pp. 49-58). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

DeCorte, W. (2000). Estimating the classification efficiency of a test battery. *Educational and Psychological Measurement, 60,* 73-85.

Drasgow, F., Embretson, S.E., Kyllonen, P.C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery (ASVAB)* (FR-06-25). Alexandria, VA: Human Resources Research Organization.

Horst, P. (1954). A technique for the development of a differential predictor battery. *Psychometrika,* 68.

Horst, P. S. (1955). A technique for the development of an absolute prediction battery. *Psychometrika*, 69.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Ingerick, M., Diaz, T., & Putka, D. (2009). *Investigations into Army enlisted classification systems: Concurrent validation report* (Technical Report 1244). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., & Heffner, T. S. (Eds.). (2010). *Expanded Enlistment Eligibility Metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Campbell, R.C. (Eds). (2006). *Army enlisted personnel competency assessment program: Phase II report* (Technical Report 1174). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Heffner, T.S. (Eds.) (2009). *Predicting Future Force Performance (Army Class): End of Training Longitudinal Validation* (Technical Report 1257). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D.J., & Owens, K.S., & Allen, M.T. (Eds.) (2010). *Validating Future Force Performance Measures (Army Class): In-Unit Performance Longitudinal Validation*. (FR 10-38). Alexandria, VA: Human Resources Research Organization.

Knapp, D.J., & Tremble, T.R. (Eds) (2007). *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McMichael, W. H. (2008, October 14). Shaky economy helps recruiting, retention. *Army Times*. Retrieved November 3, 2010, from http://www.armytimes.com/news/2008/10 /military_recruiting_2008_101008w/.

McMichael, W. H. (2009, October 15). Economy fueled recruiting gains in FY09. *Army Times*. Retrieved November 3, 2010, from http://www.armytimes.com/news/2009/10/ military_recruiting_retention_101309w/.

Moriarty, K.O., Campbell, R.C., Heffner, T.S., & Knapp, D.J. (2009). *Validating future force performance measures (Army Class): Reclassification test and criterion development* (Research Product 2009-11). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Press, W.H., Flannery, B.P., Teukolsky, S.A., & Vetterling, W.T. (1990). *Numerical recipes: The art of scientific computing*. New York: Cambridge University Press.

Putka, D. J., Le, H., McCloy, R. A., Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology, 93*, 959-981.

Putka, D.J., & Van Iddekinge, C.H. (2007). Work Preferences Survey. In D.J. Knapp & T.R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Technical Report 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3-32.

Schafer, S. M. (2007, May 4). Good economy makes recruiting tough, personnel chief says. *Associated Press*. http://www.armytimes.com/news/2007/05/ap_economyrecruit_ 070504/.

Stark, S. (2002). *A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment* [Doctoral Dissertation]. University of Illinois at Urbana-Champaign.

Stark, S. E., Hulin, C. L., Drasgow, F., & Lopez-Rivas, G. (2006). *Technical report 2 for the SBIR Phase II funding round, topic A04-029: Behavior domains assessed by TAPAS* (DCG200608). Urbana, IL: Drasgow Consulting Group.

Stark, S., & Chernyshenko, O.S., & Drasgow, F. (2010a). Adaptive testing with the Multi-Unidimensional Pairwise Preference model. Manuscript submitted for publication.

Stark, S., Chernyshenko, O.S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise preference model. *Applied Psychological Measurement, 29*, 184-201.

Stark, S., Chernyshenko, O.S., & Drasgow, F. (2010b). Tailored adaptive personality assessment system (TAPAS-95s). In D.J. Knapp & T.S. Heffner (Eds.) *Expanded enlistment eligibility metrics (EEEM): Recommendations on a non-cognitive screen for new soldier selection* (Technical Report 1267). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Stark, S., Chernyshenko, O.S., & Drasgow, F. (September, 2010c).*Update on the Tailored Adaptive Personality Assessment System (TAPAS): Results and ideas to meet the challenges of high stakes testing.* Paper presented at the 52nd annual conference of the International Military Testing Association. Lucerne, Switzerland.

Stark, S., Chernyshenko, O.S., & Drasgow, F., & Williams, B.A. (2006). Item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91,* 25-39.

Strickland, W.J. (Ed.) (2005). *A longitudinal examination of first term attrition and reenlistment among FY1999 enlisted accessions* (Technical Report 1172). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Van Iddekinge, C.H., Putka, D.J., & Sager, C.E. (2005). Attitudinal criteria. In D.J. Knapp & T.R. Tremble (Eds.), *Development of experimental Army enlisted personnel selection and classification tests and job performance criteria* (pp. 89-104) (Technical Report 1168). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

White, L.A., & Young, M.C. (1998, August). *Development and validation of the Assessment of Individual Motivation (AIM).* Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

White, L.A., Young, M.C., Hunter, A.E., & Rumsey, M.G. (2008). Lessons learned in transitioning personality measures from research to operational settings. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(3), 291-295.

APPENDIX A: BIVARIATE TAPAS CORRELATION TABLES

Table A.1. TAPAS Intercorrelations for the 13-Dimension Computer-Adaptive (13D-CAT) Version (Applicant Sample)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Achievement														
2.	Adjustment	.													
3.	Attention Seeking	-.03	.												
4.	Cooperation	**.07**	.	.05											
5.	Dominance	**.34**	.	**.18**	**.07**										
6.	Even Tempered	**.13**	.	**-.07**	**.23**	-.05									
7.	Generosity	**.15**	.	-.03	**.19**	**.06**	**.16**								
8.	Intellectual Eff	**.26**	.	.03	.04	**.26**	**.08**	.00							
9.	Non-Delinquency	**.17**	.	**-.17**	**.17**	.03	**.17**	**.17**	.01						
10.	Optimism	**.13**	.	**.12**	**.13**	**.17**	**.15**	**.08**	**.12**	**.11**					
11.	Order	**.20**	.	**-.15**	**.07**	**.11**	.01	.04	.02	**.13**	.01				
12.	Physical Condition	**.15**	.	**.11**	-.05	**.16**	-.03	**-.10**	-.02	-.04	**.06**	**.11**			
13.	Self-Control	.	.	.	.	.	.	.	.	.	.	.	.		
14.	Sociability	.01	.	**.40**	**.17**	**.26**	.01	**.10**	**.08**	-.02	**.19**	**-.08**	**.09**	.	
15.	Tolerance	**.09**	.	.02	**.19**	.04	**.21**	**.33**	**.06**	**.07**	**.08**	.01	**-.07**	.	**.15**

Note. N = 1,311. Coefficients in bold are statistically significant, $p < .05$. Results are limited to the "Applicant Sample" (Non-prior service, Education Tier 1, AFQT Category IV and above). The Adjustment and Self-Control scales were not administered with the 13D-CAT Version.

Table A.2. TAPAS Intercorrelations for the 15-Dimension Static (15D-Static) Version (Applicant Sample)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Achievement														
2.	Adjustment	**.11**													
3.	Attention Seeking	**.05**	**.13**												
4.	Cooperation	**-.03**	**.08**	**.03**											
5.	Dominance	**.31**	**.15**	**.25**	**-.08**										
6.	Even Tempered	**.05**	**.17**	.01	**.16**	**-.11**									
7.	Generosity	**.08**	-.02	**-.11**	**.18**	-.02	**.08**								
8.	Intellectual Eff	**.25**	**.18**	**.08**	**-.13**	**.26**	-.02	**-.05**							
9.	Non-Delinquency	**.19**	**.02**	**-.12**	**.11**	-.01	**.15**	**.19**	-.01						
10.	Optimism	**.26**	**.35**	**.20**	**.09**	**.23**	**.14**	.01	**.08**	**.10**					
11.	Order	**.16**	**-.07**	**-.07**	-.01	**.07**	-.02	.02	.02	**.10**	**.02**				
12.	Physical Condition	**.12**	**.09**	**.09**	**-.04**	**.17**	**-.13**	**-.03**	**.07**	.00	**.13**	**.06**			
13.	Self-Control	**.15**	.00	**-.18**	**.15**	**-.09**	**.18**	**.13**	.00	**.27**	**.04**	**.17**	**-.09**		
14.	Sociability	.01	**.16**	**.36**	**.17**	**.22**	**-.02**	-.01	.00	**-.11**	**.23**	**-.03**	**.14**	**-.21**	
15.	Tolerance	**.13**	**-.02**	.02	**.12**	**.05**	**.07**	**.29**	**.08**	**.12**	**.04**	**.04**	**-.03**	**.14**	**.07**

Note. N = 8,224. Coefficients in bold are statistically significant, $p < .05$. Results are limited to the "Applicant Sample" (Non-prior service, Education Tier 1, AFQT Category IV and above).

Table A.3. TAPAS Intercorrelations for the 15-Dimension Computer-Adaptive (15D-CAT) Version (Applicant Sample)

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	Achievement														
2.	Adjustment	**.09**													
3.	Attention Seeking	**.04**	**.11**												
4.	Cooperation	**.11**	**.12**	**.06**											
5.	Dominance	**.33**	**.10**	**.20**	.00										
6.	Even Tempered	**.10**	**.19**	**-.01**	**.25**	**-.05**									
7.	Generosity	**.08**	**-.02**	**-.09**	**.19**	.01	**.11**								
8.	Intellectual Eff	**.25**	**.18**	**.08**	**.04**	**.25**	**.08**	**-.02**							
9.	Non-Delinquency	**.18**	.00	**-.13**	**.17**	**-.02**	**.19**	**.14**	**.02**						
10.	Optimism	**.19**	**.28**	**.18**	**.17**	**.16**	**.18**	**.04**	**.10**	**.08**					
11.	Order	**.15**	**-.08**	**-.10**	.00	**.04**	**-.03**	**.04**	**.01**	**.09**	**-.02**				
12.	Physical Condition	**.15**	**.06**	**.12**	-.01	**.18**	**-.07**	**-.04**	**.04**	**-.03**	**.10**	**.02**			
13.	Self-Control	**.22**	**.06**	**-.12**	**.11**	**.05**	**.19**	**.08**	**.18**	**.23**	**.05**	**.18**	**-.05**		
14.	Sociability	**.05**	**.11**	**.36**	**.19**	**.22**	**.04**	**.07**	.00	**-.04**	**.24**	**-.05**	**.13**	**-.11**	
15.	Tolerance	**.11**	**.02**	**.02**	**.15**	**.06**	**.13**	**.32**	**.07**	**.06**	**.09**	**.03**	**-.06**	**.11**	**.11**

Note. N = 42,130. Coefficients in bold are statistically significant, $p < .05$. Results are limited to the "Applicant Sample" (Non-prior service, Education Tier 1, AFQT Category IV and above).

Table A.4. TAPAS-95s Intercorrelations from the Expanded Enlistment Eligibility Metrics (EEEM) Research

		1	2	3	4	5	6	7	8	9
1.	Achievement									
2.	Attention Seeking	**-.12**								
3.	Cooperation	-.01	**-.06**							
4.	Dominance	**.13**	**.14**	**-.13**						
5.	Even Tempered	**.05**	**-.12**	**.14**	**-.06**					
6.	Intellectual Eff	**.16**	**-.08**	**-.08**	**.15**	**.15**				
7.	Non-Delinquency	**.16**	**-.37**	**.20**	.00	**.11**	.03			
8.	Order	**.17**	**-.08**	.02	**.06**	-.01	**.07**	**.14**		
9.	Physical Condition	**.18**	**.11**	**-.13**	**.05**	-.01	.02	**-.11**	**.05**	
10.	Tolerance	**.06**	**-.04**	-.03	**.10**	**.07**	**.14**	**.05**	**.07**	.00

Note. *N* = 3,381. Coefficients in bold are statistically significant, p < .05. Results are limited to the Education Tier 1 non-prior service Soldiers.

Table A.5. TAPAS Intercorrelations for the 13-Dimension Computer-Adaptive (13D-CAT) Version (Accession Sample)

		1	2	3	4	5	6	7	8	9
1.	Achievement									
2.	Attention Seeking	-.01								
3.	Cooperation	.04	.06							
4.	Dominance	**.36**	**.20**	.04						
5.	Even Tempered	**.11**	-.07	**.22**	-.07					
6.	Intellectual Eff	**.24**	.04	.02	**.27**	.05				
7.	Non-Delinquency	**.13**	**-.20**	**.16**	-.02	**.13**	-.03			
8.	Order	**.20**	**-.17**	.04	**.10**	.05	.05	**.14**		
9.	Physical Condition	**.18**	**.12**	**-.09**	**.18**	-.06	.02	-.07	.09	
10.	Tolerance	**.11**	.03	**.15**	**.09**	**.15**	.06	.05	.05	-.06

Note. *N* = 786. Coefficients in bold are statistically significant, *p* < .05. Results are limited to the "Accession Sample" (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract).

Table A.6. TAPAS Intercorrelations for the 15-Dimension Static (15D-Static) Version (Accession Sample)

		1	2	3	4	5	6	7	8	9
1.	Achievement									
2.	Attention Seeking	**.06**								
3.	Cooperation	**-.04**	**.04**							
4.	Dominance	**.32**	**.26**	**-.08**						
5.	Even Tempered	**.06**	.01	**.16**	**-.11**					
6.	Intellectual Eff	**.23**	**.05**	**-.12**	**.24**	-.02				
7.	Non-Delinquency	**.19**	**-.11**	**.11**	.00	**.15**	-.01			
8.	Order	**.17**	**-.07**	-.01	**.08**	.00	**.04**	**.12**		
9.	Physical Condition	**.14**	**.11**	**-.03**	**.18**	**-.11**	**.10**	.00	**.07**	
10.	Tolerance	**.13**	.01	**.12**	**.04**	**.07**	**.08**	**.11**	**.04**	-.02

Note. N = 18,217. Coefficients in bold are statistically significant, $p < .05$. Results are limited to the "Accession Sample" (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract).

Table A.7. TAPAS Intercorrelations for the 15-Dimension Computer-Adaptive (15D-CAT) Version (Accession Sample)

		1	2	3	4	5	6	7	8	9
1.	Achievement									
2.	Attention Seeking	**.04**								
3.	Cooperation	**.09**	**.05**							
4.	Dominance	**.33**	**.20**	-.01						
5.	Even Tempered	**.09**	-.01	**.24**	**-.06**					
6.	Intellectual Eff	**.24**	**.07**	**.04**	**.24**	**.07**				
7.	Non-Delinquency	**.18**	**-.13**	**.17**	**-.02**	**.18**	**.03**			
8.	Order	**.16**	**-.09**	.00	**.03**	**-.02**	**.02**	**.10**		
9.	Physical Condition	**.15**	**.11**	**-.02**	**.18**	**-.09**	**.04**	**-.04**	**.02**	
10.	Tolerance	**.10**	**.02**	**.16**	**.06**	**.13**	**.07**	**.05**	**.02**	**-.06**

Note. N = 4,258. Coefficients in bold are statistically significant, $p < .05$. Results are limited to the "Accession Sample" (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract).

APPENDIX B: COMPLETE TAPAS SUBGROUP MEAN DIFFERENCES

Table B.1. TOPS Subgroup Mean Differences for Applicant Sample

	Ethnicity					Race					Gender				
	Non-Hispanic (NH)		Hispanic (H)		NH-H	White (W)		Black (B)		W-B	Male (M)		Female (F)		M-F
Scale/Predictor	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
Standardized TAPAS Scales															
Achievement	0.01	1.01	-0.05	0.95	**0.06**	0.01	1.01	-0.05	0.95	**0.06**	0.00	1.01	0.01	0.96	-0.01
Adjustment	0.03	1.01	-0.14	0.94	**0.16**	0.02	1.01	-0.07	0.97	**0.09**	0.06	1.00	-0.23	0.98	**0.29**
Attention Seeking	0.01	1.01	-0.04	0.96	**0.05**	0.02	1.01	-0.08	0.93	**0.10**	0.02	1.00	-0.10	0.99	**0.12**
Cooperation	0.00	1.00	-0.03	0.98	**0.03**	-0.01	1.00	0.01	0.99	-0.01	0.00	1.00	0.00	1.01	0.00
Dominance	0.00	1.01	0.01	0.93	-0.02	0.01	1.02	0.04	0.89	**-0.04**	0.03	1.01	-0.10	0.97	**0.13**
Even Tempered	0.01	1.01	-0.06	0.95	**0.08**	0.00	1.01	-0.01	0.98	0.01	0.02	1.00	-0.08	1.00	**0.10**
Generosity	-0.02	1.01	0.04	0.94	**-0.06**	-0.02	1.01	0.10	0.97	**-0.12**	-0.08	0.99	0.33	0.96	**-0.41**
Intellectual Eff.	0.02	1.01	-0.14	0.93	**0.16**	0.02	1.01	-0.10	0.93	**0.12**	0.04	1.01	-0.16	0.94	**0.19**
Non-Delinquency	0.00	1.01	-0.04	0.97	**0.04**	0.00	1.01	0.04	0.98	**-0.04**	-0.03	1.01	0.12	0.96	**-0.15**
Optimism	0.00	1.01	0.01	0.95	0.00	0.01	1.01	0.03	0.96	-0.02	0.01	1.00	-0.04	1.01	**0.05**
Order	-0.03	1.01	0.14	0.96	**-0.16**	-0.04	1.01	0.19	0.97	**-0.23**	-0.03	0.99	0.12	1.04	**-0.15**
Physical Condition	0.02	1.01	-0.07	0.93	**0.09**	0.03	1.01	-0.16	0.97	**0.19**	0.08	0.99	-0.31	0.97	**0.39**
Self-Control	-0.02	1.00	0.08	0.97	**-0.10**	-0.02	1.00	0.19	0.99	**-0.21**	-0.01	1.00	0.03	1.00	**-0.03**
Sociability	0.00	1.01	0.01	0.96	-0.01	0.01	1.01	-0.07	0.95	**0.08**	0.00	1.00	0.00	0.99	0.00
Tolerance	-0.05	1.01	0.23	0.90	**-0.28**	-0.04	1.01	0.19	0.91	**-0.24**	-0.07	1.00	0.28	0.96	**-0.36**
Can-Do Composite	0.02	1.00	-0.11	0.97	**0.13**	0.02	1.00	-0.03	0.98	**0.05**	0.01	1.00	-0.05	1.00	**0.06**
Will-Do Composite	0.01	1.00	-0.08	0.97	**0.10**	0.01	1.01	-0.04	0.98	**0.05**	0.02	1.00	-0.07	1.00	**0.08**

Note. Ethnicity NH *n* = 34,079-34,824, H *n* = 6,891-6,937. Race W *n* = 33,267-33,984, B *n* = 5,385-5,512. Gender M *n* = 40,416-41,450, F *n* = 9,938-10,215. *d* = Standardized mean difference (Cohen's *d*). Results are limited to the "Applicant Sample" (Non-prior service, Education Tier 1, AFQT Category IV and above). Applicants with flagged TAPAS data were also excluded from these analyses. Coefficients in bold were statistically significant using an independent samples t-test ($p < .05$).

Table B.2. TOPS Subgroup Mean Differences for Accession Sample

	Ethnicity					Race					Gender				
	Non-Hispanic (NH)		Hispanic (H)		NH-H	White (W)		Black (B)		W-B	Male (M)		Female (F)		M-F
Scale/Predictor	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*	*M*	*SD*	*M*	*SD*	*d*
Standardized TAPAS Scales															
Achievement	0.03	1.01	-0.04	0.96	**0.07**	0.03	1.01	-0.03	0.94	**0.06**	0.01	1.01	0.02	0.95	-0.01
Adjustment	0.08	1.02	-0.09	0.95	**0.17**	0.08	1.01	-0.03	0.99	**0.11**	0.11	1.00	-0.19	0.99	**0.29**
Attention Seeking	0.02	1.01	-0.03	0.95	**0.05**	0.03	1.02	-0.07	0.91	**0.10**	0.02	1.01	-0.08	0.96	**0.10**
Cooperation	0.02	1.00	0.01	1.01	0.01	0.01	1.00	0.04	0.97	-0.03	0.02	1.00	0.03	1.00	-0.01
Dominance	-0.02	1.03	0.03	0.95	**-0.05**	-0.01	1.03	0.01	0.90	-0.01	0.01	1.02	-0.12	0.97	**0.13**
Even Tempered	0.05	1.00	-0.01	0.94	**0.06**	0.04	1.00	0.05	0.98	-0.02	0.06	0.99	-0.03	0.97	**0.09**
Generosity	-0.06	1.01	0.00	0.95	**-0.06**	-0.07	1.01	0.09	0.99	**-0.16**	-0.11	1.00	0.31	0.98	**-0.42**
Intellectual Eff	0.07	0.99	-0.08	0.93	**0.15**	0.07	1.00	-0.03	0.91	**0.10**	0.08	1.00	-0.11	0.92	**0.20**
Non-Delinquency	0.04	0.99	-0.04	0.96	**0.08**	0.04	0.99	0.08	0.97	-0.04	0.01	0.99	0.16	0.95	**-0.14**
Optimism	0.03	1.00	0.03	0.96	0.01	0.04	0.99	0.05	0.97	-0.01	0.03	0.99	0.01	1.03	0.03
Order	-0.08	1.00	0.05	0.98	**-0.13**	-0.09	1.00	0.12	0.98	**-0.21**	-0.08	0.99	0.06	1.06	**-0.14**
Physical Condition	0.03	1.01	-0.05	0.94	**0.08**	0.04	1.00	-0.14	0.96	**0.18**	0.07	1.00	-0.30	0.96	**0.38**
Self-Control	0.00	1.00	0.07	0.98	**-0.08**	-0.01	1.00	0.21	0.99	**-0.23**	0.01	1.00	0.03	1.00	-0.02
Sociability	-0.02	1.02	-0.01	0.97	-0.01	-0.01	1.02	-0.08	0.96	**0.08**	-0.02	1.01	0.01	0.99	-0.03
Tolerance	-0.06	1.01	0.23	0.91	**-0.29**	-0.06	1.01	0.19	0.93	**-0.25**	-0.07	1.01	0.29	0.96	**-0.36**
Can-Do Composite	0.08	0.99	-0.06	0.97	**0.14**	0.08	0.99	0.05	0.97	0.03	0.07	0.99	0.02	0.98	**0.06**
Will-Do Composite	0.05	1.00	-0.05	0.96	**0.10**	0.05	1.00	0.01	0.96	0.03	0.06	1.00	-0.03	0.98	**0.08**

Note. Ethnicity NH *n* = 15,323-15,786, H *n* = 2,843-2,863. Race W *n* = 14,930-15,376, B *n* = 1,979-2,040. Gender M *n* = 18,772-19,416, F *n* = 3,703-3,842. *d* = Standardized mean difference (Cohen's *d*). Results are limited to the "Accession Sample" (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Applicants with flagged TAPAS data were also excluded from these analyses. Coefficients in bold were statistically significant using an independent samples t-test ($p < .05$).

APPENDIX C: DESCRIPTIVE STATISTICS FOR THE FULL SCHOOLHOUSE SAMPLE

Table C.1. Descriptive Statistics for Training Criteria Based on the Full Schoolhouse Sample

Measure/Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
Army Life Questionnaire (ALQ)						
Affective Commitment[a]	7,700	3.82	0.69	1.00	5.00	.86
Normative Commitment[a]	7,700	4.14	0.72	1.00	5.00	.80
Career Intentions[b]	7,700	3.08	1.10	1.00	5.00	.92
Reenlistment Intentions[c]	7,700	3.56	0.99	1.00	5.50	.85
Attrition Cognition[d]	7,700	1.56	0.64	1.00	5.00	.80
Army Life Adjustment[a]	7,700	4.01	0.67	1.00	5.00	.86
Army Civilian Comparison[e]	7,700	3.81	0.80	.00	5.00	.80
MOS Fit[a]	7,700	3.77	0.86	1.00	5.00	.92
Army Fit[a]	7,700	4.01	0.61	1.00	5.00	.86
Training Achievement[f]	7,686	0.41	0.61	.00	2.00	n/a
Training Failure[f]	7,700	0.42	0.66	.00	4.00	n/a
Disciplinary Incidents[f]	3,778	0.24	0.57	.00	6.00	n/a
Last APFT Score	7,581	245.53	33.02	21.00	300.00	n/a
MOS-Specific Job Knowledge Test (JKT)						
11B/11C/11X/18X	3,019	58.64	9.21	25.00	84.78	.77
19K	12	71.17	5.87	62.00	82.00	--
31B	419	70.64	9.45	38.83	91.26	.82
68W	1,657	74.70	10.07	29.35	93.48	.86
88M	753	67.15	11.12	33.33	93.06	.78
91B	144	57.71	14.29	26.80	86.60	.91
WTBD Job Knowledge	7,433	65.21	12.67	12.90	100.00	.64
Army-Wide Performance Rating Scales[g]						
Effort	4,123	4.73	1.22	1.00	7.00	n/a
Physical Fitness & Bearing	4,138	4.65	1.22	1.00	7.00	n/a
Personal Discipline	4,187	4.81	1.22	1.00	7.00	n/a
Commitment &Adjustment	4,188	4.84	1.19	1.00	7.00	n/a
Support for Peers	4,160	4.84	1.16	1.00	7.00	n/a
Peer Leadership	3,746	4.52	1.36	1.00	7.00	n/a

Table C.1. (Continued)

Measure/Scale	*n*	*M*	*SD*	*Min*	*Max*	*α*
Common Warrior Tasks Knowledge and Skill	3,788	4.68	1.20	1.00	7.00	n/a
MOS Qualification Knowledge and Skill	3,452	4.80	1.11	1.00	7.00	n/a
Overall Performance Scale	3,994	3.41	.81	1.00	5.00	n/a
MOS-Specific Performance Rating Composite Scores						
Total (combined across MOS)	3,709	4.52	0.94	1.00	7.00	n/a
11B/11C/11X/18X	1,303	4.61	0.86	1.00	7.00	.95
31B	188	4.70	1.09	1.00	6.50	.95
68W	1,611	4.36	0.97	1.00	7.00	.94
88M	498	4.51	0.64	2.25	7.00	.90
91B	109	5.58	1.37	3.00	7.00	.98

Note. n/a = Internal consistency/coefficient alpha could not be computed for the scales/measures. Job knowledge scores are percent correct.
[a] These items were responded using agreement scales (1=Strongly Disagree, 2=Disagree, 3=Neither Agree Nor Disagree, 4=Agree, 5=Strongly Agree).
[b] This construct was measured by items using three types of scales: agreement scale (same as above), confident scale (1=Not At All Confident, 2= Somewhat Confident, 3=Confident, 4=Very Confident, 5=Extremely Confident), and likelihood scale (1=Extremely Unlikely, 2=Unlikely, 3=Neither Likely Nor Unlikely, 4=Likely, 5=Extremely Likely).
[c] This construct was measured by items using agreement scale (same as above) and likelihood scale (same as above).
[d] This construct was measured by items using agreement scale (same as above) and often scale (1=Never, 2=Rarely, 3=Sometimes, 4=Often, 5=Very Often).
[e] This construct was measured by the following scales: 1=Much Better in the Army, 2=Better in the Army, 3=About the Same, 4=Better in Civilian Life, 5=Much Better in Civilian Life.
[f] These scales are the total number of 'YES' responses to a series of yes/no questions about things that happened in training.
[g] All Performance Rating Scale scores range between 1 and 7, except for the "Overall Performance Scale," which ranges from 1 to 5 (see Figure 5.1).

Table C.2. Descriptive Statistics for Schoolhouse Criteria by MOS (Full Schoolhouse Sample)

	Total		11B		19K		25U		31B		42A		68W		88M		91B	
Measure/Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Army Life Questionnaire (ALQ)																		
Affective Commitment	3.82	0.69	3.87	0.67	3.60	0.36	3.58	0.72	4.04	0.56	3.90	0.68	3.69	0.73	3.90	0.67	3.79	0.68
Normative Commitment	4.14	0.72	4.16	0.70	3.73	0.75	3.97	0.81	4.28	0.64	4.10	0.76	4.10	0.73	4.19	0.70	4.00	0.79
Career Intentions	3.08	1.10	3.10	1.08	2.67	0.99	2.91	1.18	3.18	1.02	3.31	1.12	2.90	1.11	3.34	1.12	3.06	1.11
Reenlistment Intentions	3.56	0.99	3.56	0.97	3.27	0.69	3.36	1.06	3.66	0.93	3.66	1.05	3.43	1.02	3.81	0.97	3.59	0.93
Attrition Cognitions	1.56	0.64	1.55	0.66	1.60	0.53	1.73	0.72	1.43	0.51	1.59	0.64	1.61	0.63	1.52	0.62	1.64	0.68
Army Life Adjustment	4.01	0.67	3.99	0.69	3.94	0.47	3.98	0.62	4.10	0.60	4.02	0.70	3.96	0.65	4.08	0.62	3.92	0.72
Army Civilian Comparison	3.81	0.80	3.88	0.79	3.92	0.36	3.72	0.78	4.00	0.71	4.00	0.72	3.58	0.81	3.96	0.79	3.94	0.78
MOS Fit	3.77	0.86	3.79	0.84	2.94	0.93	3.31	0.83	4.02	0.74	3.68	0.90	3.98	0.81	3.33	0.85	3.70	0.89
Army Fit	4.01	0.61	4.04	0.60	3.88	0.34	3.84	0.63	4.20	0.51	4.14	0.60	3.89	0.62	4.12	0.60	3.96	0.63
Training Achievement	0.41	0.61	0.48	0.67	0.42	0.79	0.46	0.60	0.31	0.55	0.42	0.61	0.28	0.46	0.35	0.54	0.47	0.60
Training Failure	0.42	0.66	0.28	0.53	0.33	0.65	0.73	0.82	0.30	0.54	0.70	0.80	0.59	0.75	0.49	0.68	0.59	0.75
Disciplinary Incidents	0.24	0.57	0.25	0.58	0.50	0.67	0.00	0.00	1.00	--	--	--	0.33	0.58	0.28	0.52	0.80	1.30
Last APFT Score	245.53	33.02	241.72	33.15	252.00	24.06	243.72	34.90	256.71	31.75	248.97	29.93	248.95	31.89	242.98	32.55	243.85	27.42
MOS-Specific JKT	64.98	9.82	58.64	9.21	71.17	5.87	--	--	70.64	9.45	--	--	74.70	10.07	67.15	11.12	57.71	14.29
WTBD JKT	65.21	12.67	64.75	12.72	77.15	8.96	58.60	12.50	69.08	10.75	60.66	12.69	68.14	11.54	62.65	12.65	58.08	12.13
Army-Wide PRS																		
Effort	4.73	1.22	4.71	1.26	--	--	4.59	0.94	4.82	1.41	--	--	4.75	1.21	4.64	0.99	4.96	1.43
Physical Fitness & Bearing	4.65	1.22	4.61	1.22	--	--	4.61	1.02	4.78	1.44	--	--	4.67	1.24	4.49	0.96	5.09	1.54
Personal Discipline	4.81	1.22	4.86	1.22	--	--	4.72	1.05	4.88	1.58	--	--	4.78	1.25	4.57	0.95	5.12	1.46
Commitment & Adjustment	4.84	1.19	4.91	1.19	--	--	4.71	1.05	5.06	1.34	--	--	4.70	1.22	4.74	0.89	5.40	1.52
Support for Peers	4.84	1.16	4.86	1.20	--	--	4.88	0.92	5.05	1.35	--	--	4.76	1.15	4.73	0.90	5.38	1.56
Peer Leadership	4.52	1.36	4.38	1.34	--	--	4.67	1.29	4.70	1.62	--	--	4.53	1.43	4.49	0.95	5.22	1.51
Common Warrior Tasks KS	4.68	1.20	4.67	1.07	--	--	4.83	1.04	5.16	1.23	--	--	4.56	1.46	4.53	0.81	5.18	1.56
MOS Qualification KS	4.80	1.11	4.74	1.08	--	--	4.82	1.05	4.98	1.38	--	--	4.87	1.19	4.63	0.88	5.46	1.36
Overall Performance	3.41	0.81	3.38	0.80	--	--	3.28	0.65	3.38	0.97	--	--	3.39	0.87	3.49	0.61	3.71	0.85
MOS-Specific Performance Composite	4.52	0.94	4.60	0.87	--	--	--	--	4.70	1.09	--	--	4.35	0.97	4.51	0.64	5.58	1.37

Note. The analyses were conducted using the full schoolhouse dataset. Job knowledge test scores are percent correct. Due to low sample size, the AW PRS were not computed for 19K and 42A. KS = Knowledge and Skills. Sample sizes can be found in Table C.1.

Table C.3 Interrater Reliability Estimates for the Army-Wide and MOS-Specific PRS using the Full Schoolhouse Sample

PRS Scales	Total	11B	25U	31B	68W	88M	91B
Army-Wide PRS							
Effort	.17	.30	.01	.46	.03	.05	.51
Physical Fitness & Bearing	.24	.32	.13	.30	.34	.08	.40
Personal Discipline	.17	.27	.06	.43	.30	.01	.45
Commitment & Adjustment	.14	.22	.07	.33	.28	.00	.38
Support for Peers	.15	.27	.00	.21	.13	.04	.26
Peer Leadership	.22	.33	.11	.27	.21	.06	.34
Common Warrior Tasks KS	.08	.19	.07	.20	.02	.00	.34
MOS Qualification KS	.10	.24	.00	.14	.00	.05	.20
Overall Performance	.23	.39	.05	.34	.05	.12	.69
Avg. MOS-specific PRS	--	.13	--	.09	.00	.00	.12

Note. Because the measurement design was ill-structured, interrater reliability was estimated using G(q,k) (Putka, Le, McCloy, & Diaz, 2008). Avg. MOS-specific PRS = The average G(q,k) estimate across the MOS-specific scales for the target MOS; MOS-specific scales were not administered to 25U. The number of raters providing ratings (AW PRS k = 331-368; 11B k = 127-131; 25U k = 19-22; 31B k = 13; 68W k = 72-112; 88M k = 85-91; 91B k = 6) and number of targets rated (AW PRS n = 3,707-4,128; 11B n = 1,211-1,687; 25U n = 207-233; 31B n = 179-182; 68W n = 1,071-1,405; 88M k = 450-487; 91B k = 100) varied by MOS. Coefficients for 19K (n = 0) and 42A (n = 8 with ratings) were not computed due to insufficient sample size.

Table C.4. Army Life Questionnaire (ALQ) Intercorrelations for the Full Schoolhouse Sample

Scale	1	2	3	4	5	6	7	8	9	10	11	12
1. Affective Commitment												
2. Normative Commitment	**.68**											
3. Career Intentions	**.59**	**.44**										
4. Reenlistment Intentions	**.56**	**.46**	**.85**									
5. Attrition Cognition	**-.63**	**-.74**	**-.47**	**-.49**								
6. Army Life Adjustment	**.45**	**.47**	**.36**	**.40**	**-.57**							
7. Army Civilian Comparison	**.42**	**.31**	**.33**	**.34**	**-.33**	**.22**						
8. General MOS Fit	**.54**	**.48**	**.31**	**.31**	**-.49**	**.38**	**.25**					
9. Army Fit	**.83**	**.71**	**.56**	**.55**	**-.69**	**.61**	**.43**	**.56**				
10. Training Achievement	**.07**	.02	**.09**	**.07**	**-.06**	**.14**	**.02**	**.07**	**.09**			
11. Training Restart	**-.08**	**-.09**	**-.03**	**-.04**	**.15**	**-.22**	.00	**-.10**	**-.11**	**-.12**		
12. Disciplinary Action	**-.08**	**-.11**	**-.05**	**-.07**	**.16**	**-.21**	-.02	**-.11**	**-.13**	**-.06**	**.18**	
13. Army Physical Fitness Test	**.06**	**.09**	**.03**	**.04**	**-.14**	**.25**	**-.03**	**.11**	**.12**	**.23**	**-.26**	**-.17**

Note. Significant correlation coefficients are bolded (p < .05). Sample sizes for each research criterion variable can be found in Table C.1.

Table C.5. MOS Job Knowledge Test (JKT) Correlations with the WTBD JKT in Full Schoolhouse Sample

Measure/Scale	WTBD JKT
MOS-Specific Job Knowledge Test (JKT)	
1. MOS z Scores	**.54**
2. 11B/11C/11X/18X	**.18**
3. 19K	.20
4. 31B	**.48**
5. 68W	**.48**
6. 88M	**.23**
7. 91B	**.27**

Note. Significant correlation coefficients are bolded ($p < .05$). MOS z scores = MOS-Specific JKTs standardized and combined into one variable. Sample sizes for each research criterion variable can be found in Table C.1.

Table C.6. Army-Wide and MOS-Specific Performance Rating Scale (PRS) Intercorrelations for the Full Schoolhouse Sample

Measure/Scale	1	2	3	4	5	6	7	8	9
Army-Wide Performance Rating Scales									
1. Effort									
2. Physical Fitness and Bearing	.73								
3. Personal Discipline	.73	.72							
4. Commitment and Adjustment	.73	.73	.77						
5. Support for Peers	.70	.67	.75	.78					
6. Peer Leadership	.66	.65	.66	.72	.74				
7. Common/Warrior Tasks Knowledge and Skills	.62	.62	.63	.71	.71	.77			
8. MOS Proficiency	.63	.66	.62	.72	.71	.70	.79		
9. Overall Performance	.59	.58	.59	.61	.58	.58	.55	.55	
MOS-Specific Performance Rating Composites									
10. Combined PRS Composites	.51	.49	.47	.54	.55	.59	.67	.72	.47
11. 11B/11C/11X/18X	.64	.66	.65	.69	.76	.69	.71	.74	.56
12. 31B	.67	.63	.65	.62	.62	.52	.69	.74	.65
13. 68W	.44	.37	.35	.44	.41	.53	.64	.63	.43
14. 88M	.57	.60	.55	.57	.60	.65	.68	.76	.50
15. 91B	.63	.74	.77	.76	.88	.86	.80	.89	.40

Note. All correlation coefficients are significant ($p < .05$). Sample sizes for each research criterion variable can be found in Table C.1.

Table C.7 Correlations between the Army Life Questionnaire (ALQ) and Job Knowledge Test (JKT) Scores for the Full Schoolhouse Sample

Army Life Questionnaire (ALQ) Scales	*MOS-Specific Job Knowledge Test (JKT)*							
	Combined	11B	19K	31B	68W	88M	91B	WTBD JKT
Affective Commitment	**.09**	**.10**	.34	**.13**	**.05**	**.08**	.04	**.08**
Normative Commitment	**.17**	**.08**	.30	**.14**	**.13**	**.10**	.09	**.18**
Career Intentions	.02	**.04**	-.11	.00	-.02	.06	.06	.01
Reenlistment Intentions	**.05**	**.04**	.15	.06	.03	**.10**	.06	**.06**
Attrition Cognition	**-.15**	**-.05**	-.15	**-.16**	**-.12**	**-.13**	-.12	**-.17**
Army Life Adjustment	**.09**	**-.04**	-.33	**.14**	**.09**	.04	**.18**	**.16**
Army Civilian Comparison	**.03**	**.11**	.01	-.07	**.09**	**.09**	.08	-.01
General MOS Fit	**.13**	.03	.32	**.11**	**.12**	-.04	.13	**.14**
Army Fit	**.13**	**.07**	.09	**.15**	**.09**	**.08**	.07	**.13**
Training Achievement	**-.13**	**-.12**	-.04	**-.11**	**-.05**	**-.08**	-.11	**-.09**
Training Restart	**-.08**	**.04**	.17	**-.16**	**-.09**	-.05	-.12	**-.14**
Disciplinary Action	**-.04**	.02	.07	--	--	--	--	**-.09**
Army Physical Fitness Test	-.01	**-.14**	.25	**.15**	-.01	**-.15**	.05	**.08**

Note. Significant correlation coefficients are bolded ($p < .05$). Sample sizes for each research criterion variable can be found in Table C.1.

Table C.8. Correlations between the Army Life Questionnaire (ALQ) and Performance Rating Scales (PRS) Scores for the Full Schoolhouse Sample

	Army Life Questionnaire (ALQ) Scales												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Army-wide Performance Rating Scales													
Effort	**.07**	**.08**	**.06**	**.07**	**-.10**	**.12**	-.03	**.09**	**.09**	**.11**	**-.13**	**-.20**	**.17**
Physical Fitness and Bearing	**.05**	**.07**	**.05**	**.06**	**-.11**	**.15**	-.03	**.10**	**.08**	**.13**	**-.14**	**-.20**	**.23**
Personal Discipline	**.07**	**.09**	**.06**	**.06**	**-.11**	**.12**	-.02	**.10**	**.09**	**.08**	**-.10**	**-.19**	**.12**
Commitment and Adjustment	**.07**	**.06**	**.07**	**.07**	**-.10**	**.13**	.00	**.08**	**.09**	**.11**	**-.11**	**-.13**	**.14**
Support for Peers	**.05**	**.06**	**.05**	**.05**	**-.09**	**.11**	-.01	**.06**	**.06**	**.10**	**-.11**	**-.20**	**.13**
Peer Leadership	**.05**	**.05**	**.05**	**.06**	**-.08**	**.12**	-.02	**.08**	**.07**	**.11**	**-.06**	**-.15**	**.15**
Common/Warrior Tasks Knowledge and Skills	.03	.03	**.04**	**.05**	**-.07**	**.10**	-.01	**.08**	**.05**	**.06**	**-.08**	**-.14**	**.12**
MOS Proficiency	**.04**	.04	.04	**.05**	**-.06**	**.12**	-.01	**.11**	**.04**	**.07**	**-.09**	**-.16**	**.14**
Overall Performance	**.07**	**.07**	**.05**	**.06**	**-.11**	**.17**	.00	**.09**	**.10**	**.13**	**-.16**	**-.25**	**.21**
MOS-Specific Performance Rating Composites													
Total	**.04**	.01	.03	.03	**-.05**	**.08**	**.05**	**.04**	**.04**	**.10**	**-.09**	**-.13**	**.07**
11B/11C/11X/18X	**.10**	**.12**	**.13**	**.14**	**-.13**	**.14**	-.01	.09	.08	**.13**	**-.14**	**-.13**	**.26**
31B	.11	.08	-.04	.02	-.13	**.23**	.05	.14	.14	**.17**	.01	--	**.24**
68W	-.01	-.02	.01	-.01	-.03	.05	.00	.05	.00	**.07**	**-.09**	--	.03
88M	.04	.01	.07	.07	-.01	**.09**	.02	.05	.01	.08	-.03	--	**.11**
91B	-.13	-.13	-.13	-.10	.11	-.06	-.02	.06	-.18	.05	.00	--	-.03

Note. Significant correlation coefficients are bolded ($p < .05$). Sample sizes for each research criterion variable can be found in Table C.1. 1=Affective Commitment; 2=Normative Commitment; 3=Career Intentions; 4=Reenlistment Intentions; 5=Attrition Cognition; 6=Army Life Adjustment; 7=Army Civilian Comparison; 8=General MOS Fit; 9=Needs Supplies Army Fit; 10=Training Achievement; 11=Training Restart; 12=Disciplinary Action; 13=Army Physical Fitness Test.

Table C.9. Correlations between Job Knowledge Test (JKT) and Performance Rating Scale (PRS) Scores for the Full Schoolhouse Sample

	MOS-Specific JKTs						
	Combined	11B	31B	68W	88M	91B	WTBD JKT
Army-Wide Performance Rating Scales							
Effort	**.05**	**-.23**	.13	.02	-.02	-.02	**.07**
Physical Fitness and Bearing	.02	**-.20**	**.15**	.00	-.05	-.14	**.06**
Personal Discipline	**.05**	**-.13**	.12	.00	.05	-.02	**.07**
Commitment and Adjustment	.01	**-.15**	.11	-.01	.04	-.08	.03
Support for Peers	-.01	**-.13**	.10	**-.06**	-.02	-.10	.03
Peer Leadership	-.01	**-.22**	.02	-.05	.02	-.03	.02
Common/Warrior Tasks Knowledge and Skills	-.02	**-.17**	.04	**-.08**	-.07	-.14	.03
MOS Proficiency	.03	**-.17**	.10	.02	-.08	-.05	**.07**
Overall Performance	.02	**-.17**	.07	-.01	-.07	.03	**.05**
MOS-Specific Performance Ratings Composite							
Combined	-.04	**-.18**	.06	**-.06**	-.08	-.04	-.01
11B/11C/11X/18X	-.06	**-.18**	--	--	--	--	**.13**
31B	.14	--	.06	--	--	--	.00
68W	**-.07**	--	--	**-.06**	--	--	.02
88M	.01	--	--	--	-.08	--	.02
91B	-.04	--	--	--	--	-.04	-.08

Note. Significant correlation coefficients are bolded ($p < .05$). Sample sizes for each research criterion variable can be found in Table C.1.

Table C.10. Descriptive Statistics for Administrative Criteria Based on the Applicant Sample by MOS

	11B/11C/11X/18X			*19K*			*31B*			*68W*			*88M*			*91B*		
Administrative Criterion	N [b]	N_{Attrit}	$\%_{Attrit}$	N [b]	N_{Attrit}	$\%_{Attrit}$	N [b]	N_{Attrit}	$\%_{Attrit}$	N [b]	N_{Attrit}	$\%_{Attrit}$	N [b]	N_{Attrit}	$\%_{Attrit}$	N [b]	N_{Attrit}	$\%_{Attrit}$
Three-Month Attrition[a]	706	42	5.9	71	5	7.0	63	4	6.3	160	4	2.5	84	5	6.0	82	5	6.1
Initial Military Training (IMT) Criteria	N [c]	$N_{Restart}$	$\%_{Restart}$	N [c]	$N_{Restart}$	$\%_{Restart}$	N [c]	$N_{Restart}$	$\%_{Restart}$	N [c]	$N_{Restart}$	$\%_{Restart}$	N [c]	$N_{Restart}$	$\%_{Restart}$	N [c]	$N_{Restart}$	$\%_{Restart}$
Restarted at Least Once During IMT	314	71	22.6	25	8	32.0	41	11	26.8	9	4	44.4	103	10	9.7	24	4	16.7
Restarted at Least Once During IMT for Pejorative Reasons	313	70	22.4	23	6	26.1	41	11	26.8	9	4	44.4	99	6	6.1	22	2	9.1
Restarted at Least Once During IMT for Academic Reasons	280	37	13.2	22	5	22.7	35	5	14.3	9	4	44.4	103	10	9.7	24	4	16.7
AIT School Grades	N [d]	*M*	*SD*	N [d]	*M*	*SD*	N [d]	*M*	*SD*	N [d]	*M*	*SD*	N [d]	*M*	*SD*	N [d]	*M*	*SD*
Overall Average (Unstandardized)	--	--	--	--	--	--	--	--	--	81	86.84	6.96	--	--	--	--	--	--
Overall Average (Standardized within MOS)	--	--	--	--	--	--	--	--	--	81	0.07	0.96	--	--	--	--	--	--

Note. Results are limited to the Applicant sample (non-prior service, Education Tier 1, AFQT Category IV or higher).

[a] Attrition results reflect Regular Army Soldiers only.

[b] N = number of Soldiers with 3-month attrition data at the time data were extracted. N_{Attrit} = number of Soldiers who attrited through 3 months of service. $\%_{Attrit}$ = percentage of Soldiers who attrited through 3 months of service [(N_{Attrit} /N) x 100].

[c] N = number of Soldiers with IMT data at the time data were extracted. $N_{Restart}$ = number of Soldiers who restarted at least once during IMT. $\%_{Restart}$ = percentage of Soldiers who restarted at least once during IMT [($N_{Restart}$ /N) x 100].

[d] N = number of Soldiers with AIT school grade data. Standardized school grades were not computed for MOS with insufficient sample size (n < 15).

APPENDIX D: SUPPLEMENTAL VALIDITY AND CLASSIFICATION TABLES

Table D.1. Incremental Validity Estimates for the TAPAS Scales over the AFQT for Predicting Performance- and Retention-related Criteria

Criterion	*n*	AFQT Only *R* (r_{pb})	AFQT + TAPAS *R* (r_{pb})	Δ*R* (Δr_{pb})
Can-do Performance				
WTBD JKT	255	**.43**	**.51**	.08
MOS-Specific JKT	203	**.31**	**.41**	.09
MOS Proficiency (PRS)	118	.04	.20	.16
MOS-Specific PRS	113	.16	.41	.25
IMT Exam Grade	544	**.23**	**.27**	.04
# of Restarts in IMT (ALQ)	670	.02	.17	.15
Graduated IMT without Restart	670	(.02)	(.22)	(.20)
Training Achievement (ALQ)	272	**.13**	**.34**	**.20**
Training Failure (ALQ)	272	.08	.29	.20
Common/Warrior Tasks KS (PRS)	123	.04	.30	.26
Will-do Performance				
Exhibiting Effort (PRS)	127	.07	.33	.26
Support for Peers (PRS)	126	.03	.31	.29
Peer Leadership (PRS)	118	.03	.33	.29
Exhibiting Fitness and Bearing (PRS)	126	.11	.43	.32
Personal Discipline (PRS)	127	.09	.38	.29
Last APFT Score (ALQ)	269	.03	**.34**	**.31**
Disciplinary Action (ALQ)	129	.05	.30	.25
Commitment and Adjustment (PRS)	127	.07	.32	.25
Retention				
Adjustment to Army Life (ALQ)	272	**.16**	**.32**	.16
Affective Commitment (ALQ)	272	.00	.22	.21
Normative Commitment (ALQ)	272	.09	.22	.12
Career Intentions (ALQ)	272	.03	.18	.16
Attrition Cognitions (ALQ)	272	.07	.23	.17
Reenlistment Intentions (ALQ)	272	.02	.20	.18
Army Fit (ALQ)	272	.06	.19	.14
MOS Fit (ALQ)	272	**.13**	.27	.15
Army Civilian Comparison (ALQ)	272	**.12**	.31	.18
3-Month Attrition[a]	2,443	(.01)	(**.09**)	(.08)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire, PRS = Performance Rating Scales. AFQT Only = Correlation between the AFQT and the criterion of interest. AFQT + TAPAS = Multiple correlation (*R*) between the AFQT and the selected predictor measure with the criterion of interest. Δ*R* = Increment in *R* over the AFQT from adding the selected predictor measure to the regression model ([AFQT + TAPAS] – AFQT Only). *Point-biserial correlation (r_{pb})* = Observed point-biserial correlation between Soldiers' predicted probability of attriting/graduating and their actual attrition/graduation behavior. Large, positive r_{pb} values mean that the TOPS composite or scale performed well in predicting actual attrition/graduation. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a]Attrition results include Regular Army Soldiers only.

Table D.2. Bivariate and Semi-Partial Correlations between the TAPAS Scales and Can-do Performance-related Criteria

	Criteria									
	WTBD JKT	MOS-Specific JKT	MOS Proficiency (PRS)	MOS-Specific PRS	IMT Exam Grade	# of Restarts in IMT (ALQ)	Graduated IMT without Restart	Training Achievement (ALQ)	Training Failure (ALQ)	Common/ Warrior Tasks KS (PRS)
TAPAS Dimensions	*N* = 342	*N* = 274	*N* = 161	*N* = 163	*N* = 660	*N* = 1,050	*N* = 1,050	*N* = 361	*N* = 361	*N* = 170
Achievement	.04 (.00)	-.04 (-.06)	.05 (.05)	-.06 (-.05)	.01 (-.02)	-.01 (-.01)	.01 (.01)	.10 **(.11)**	**-.15 (-.14)**	.06 (.06)
Adjustment[a]	.12 (.07)	.08 (.04)	.05 (.05)	-.04 (-.02)	.01 (-.01)	**.09 (.09)**	-.02 (-.02)	.09 (.11)	-.11 (-.10)	.15 (.14)
Attention Seeking	-.04 (-.08)	-.05 (-.08)	.04 (.04)	-.03 (-.01)	.00 (-.02)	.00 (.00)	.03 (.03)	**.15 (.16)**	-.05 (-.04)	.06 (.05)
Cooperation	-.06 (-.06)	.02 (.02)	-.01 (-.01)	.10 (.09)	.03 (.04)	.03 (.03)	-.05 (-.05)	-.04 (-.04)	.04 (.04)	.11 (.11)
Dominance	.02 (-.02)	-.11 **(-.14)**	-.07 (-.07)	**-.21 (-.19)**	.02 (.00)	-.04 (-.04)	**.07 (.07)**	**.15 (.16)**	-.08 (-.07)	-.10 (-.11)
Even Tempered	-.11 **(-.14)**	-.04 (-.06)	.02 (.02)	.08 (.10)	.01 (-.01)	.00 (.00)	-.02 (-.02)	.00 (.01)	.05 (.06)	.10 (.09)
Generosity	**-.17 (-.14)**	**-.18 (-.16)**	-.07 (-.07)	-.11 (-.12)	.01 (.03)	-.03 (-.03)	.04 (.04)	-.07 (-.08)	.07 (.07)	-.10 (-.09)
Intellectual Efficiency	**.20** (.01)	.11 (-.03)	.05 (.04)	-.19 (-.14)	**.11** (.01)	.05 (.04)	.00 (-.01)	-.01 (.06)	-.07 (-.03)	.06 (.04)
Non-delinquency	-.08 (-.08)	-.09 (-.08)	-.01 (-.01)	-.01 (-.01)	-.03 (-.03)	.00 (.00)	-.02 (-.02)	.08 (.08)	.04 (.04)	.08 (.08)
Optimism	.03 (.03)	-.03 (-.03)	-.05 (-.05)	-.13 (-.12)	.01 (.01)	-.02 (-.02)	.02 (.02)	.04 (.04)	-.10 (-.10)	.06 (.05)
Order	**-.13** (-.06)	-.08 (-.02)	-.02 (-.01)	-.08 (-.11)	.02 (.07)	-.07 **(-.07)**	**.08 (.08)**	**.11** (.09)	.00 (-.02)	-.02 (-.02)
Physical Conditioning	.03 (.01)	.00 (-.01)	.05 (.05)	-.03 (-.02)	-.04 (-.05)	**-.07 (-.07)**	**.10 (.10)**	**.14 (.15)**	**-.17 (-.17)**	.05 (.04)
Self-Control[a]	-.12 **(-.11)**	-.07 (-.06)	.01 (.01)	.11 (.11)	.04 (.05)	.06 (.06)	-.01 (-.01)	-.03 (-.03)	.07 (.07)	.01 (.01)
Sociability	-.03 (.00)	-.07 (-.05)	-.10 (-.09)	-.05 (-.06)	**-.09 (-.08)**	-.03 (-.02)	.05 (.05)	**.11** (.10)	.00 (-.01)	-.07 (-.07)
Tolerance	-.09 (-.08)	-.09 (-.08)	-.02 (-.02)	**-.16 (-.16)**	.01 (.02)	-.04 (-.04)	.00 (.00)	.01 (.01)	.09 (.09)	.00 (.00)
TAPAS Composites										
Can-do Composite	.03 (-.07)	-.03 (-.10)	.03 (.02)	-.11 (-.08)	.04 (-.01)	.01 (.00)	.00 (.00)	.07 **(.11)**	-.08 (-.06)	.12 (.12)
Will-do Composite	-.04 (-.05)	-.04 (-.06)	.03 (.03)	.01 (.01)	-.03 (-.03)	-.03 (-.03)	.02 (.02)	.07 (.08)	-.08 (-.08)	.09 (.09)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. PRS = Performance Ratings Scales. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in parentheses are semi-partial correlations between the TAPAS scales and the criterion of interest, controlling for AFQT. Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a] Adjustment and Self Control were included in the TAPAS 15-dimension versions (i.e., static and CAT) only. Sample sizes for these scales are smaller, ranging from 113 – 1,050.

Table D.3. Bivariate and Semi-partial Correlations between the TAPAS Scales and Will-do Performance-related Criteria

	Criteria																			
	Exhibiting Effort (PRS)		Support for Peers (PRS)		Peer Leadership (PRS)		Exhibiting Fitness & Bearing (PRS)		Personal Discipline (PRS)		Last APFT Score (ALQ)		Disciplinary Action (ALQ)		Commitment & Adjustment (PRS)	
TAPAS Dimensions	*N* = 174		*N* = 175		*N* = 165		*N* = 175		*N* = 176		*N* = 357		*N* = 176		*N* = 176	
Achievement	.05	(.04)	.04	(.04)	.05	(.05)	.15	(.14)	.13	(.12)	.05	(.05)	**-.21**	**(-.20)**	.13	(.13)
Adjustment[a]	-.02	(-.03)	-.02	(-.02)	.03	(.02)	.06	(.05)	.08	(.07)	.03	(.03)	-.08	(-.08)	.04	(.03)
Attention Seeking	-.03	(-.04)	.01	(.01)	.09	(.08)	.04	(.03)	.03	(.02)	.00	(.00)	.05	(.06)	-.05	(-.06)
Cooperation	.08	(.08)	.14	(.14)	.14	(.14)	.09	(.09)	-.05	(-.05)	.06	(.06)	.03	(.03)	-.03	(-.03)
Dominance	-.12	(-.13)	**-.15**	**(-.16)**	**-.17**	**(-.18)**	-.13	(-.15)	**-.16**	**(-.16)**	.08	(.08)	-.03	(-.03)	-.13	(-.13)
Even Tempered	**.18**	**(.17)**	.10	(.10)	.08	(.08)	.15	(.14)	.08	(.08)	-.10	(-.10)	-.04	(-.04)	.04	(.03)
Generosity	-.09	(-.08)	-.12	(-.12)	-.06	(-.06)	**-.16**	**(-.15)**	-.07	(-.06)	.06	(.06)	-.04	(-.04)	-.11	(-.11)
Intellectual Efficiency	.07	(.04)	-.03	(-.05)	.00	(-.01)	.06	(.01)	.04	(.00)	.00	(-.01)	.00	(.03)	.06	(.03)
Non-delinquency	-.04	(-.04)	.08	(.08)	.08	(.08)	.10	(.11)	.02	(.02)	.08	(.08)	-.09	(-.09)	.04	(.04)
Optimism	-.05	(-.05)	-.02	(-.02)	.00	(.00)	-.03	(-.03)	.05	(.05)	.03	(.03)	-.06	(-.06)	-.01	(-.01)
Order	-.02	(-.01)	.00	(.01)	-.02	(-.02)	-.07	(-.05)	.04	(.06)	.03	(.03)	-.01	(-.02)	-.01	(.00)
Physical Conditioning	.09	(.08)	.02	(.02)	.07	(.07)	.14	(.14)	.02	(.02)	**.27**	**(.27)**	-.10	(-.10)	.10	(.10)
Self-Control[a]	.11	(.11)	.02	(.02)	-.02	(-.02)	-.03	(-.03)	.02	(.02)	.07	(.07)	.07	(.07)	.11	(.11)
Sociability	-.05	(-.05)	-.07	(-.07)	-.05	(-.04)	-.11	(-.10)	**-.22**	**(-.22)**	.07	(.07)	.04	(.03)	-.13	(-.12)
Tolerance	.01	(.01)	.04	(.04)	.04	(.04)	-.01	(-.01)	.00	(.01)	.09	(.09)	-.05	(-.05)	.02	(.02)
TAPAS Composites																
Can-do Composite	.07	(.06)	.05	(.04)	.07	(.06)	.14	(.12)	.10	(.09)	.02	(.02)	-.14	(-.14)	.08	(.07)
Will-do Composite	.12	(.12)	.08	(.08)	.07	(.07)	**.20**	**(.19)**	.09	(.08)	**.12**	(.12)	**-.19**	**(-.19)**	.14	(.14)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. PRS = Performance Ratings Scales. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in parentheses are semi-partial correlations between the TAPAS scales and the criterion of interest, controlling for AFQT. Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a] Adjustment and Self Control were included in the TAPAS 15-dimension versions (i.e., static and CAT) only. Sample sizes for these scales are smaller, ranging from 118 – 357.

Table D.4. Bivariate and Semi-partial Correlations between the TAPAS Scales and Retention-related Criteria

	Criteria									
	Adjustment to Army Life (ALQ)	Affective Commitment (ALQ)	Normative Commitment (ALQ)	Career Intentions (ALQ)	Attrition Cognitions (ALQ)	Reenlistment Intentions (ALQ)	Army Fit (ALQ)	MOS Fit (ALQ)	Army Civilian Comparison (ALQ)	3-Month Attrition[b]
TAPAS Dimensions	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 361	*N* = 2,810
Achievement	**.13 (.12)**	.10 (.10)	**.13 (.12)**	.08 (.09)	**-.15 (-.14)**	.10 (.10)	**.11 (.11)**	**.14** (.13)	-.05 (-.04)	.01 (.01)
Adjustment[a]	**.18 (.17)**	-.04 (-.04)	.07 (.06)	.01 (.01)	-.08 (-.07)	-.02 (-.01)	-.01 (-.02)	.03 (.02)	-.09 (-.08)	.01 (.01)
Attention Seeking	.00 (-.01)	.02 (.02)	-.03 (-.04)	-.06 (-.06)	.00 (.01)	-.04 (-.03)	.03 (.02)	.03 (.02)	.06 (.07)	.00 (.00)
Cooperation	-.02 (-.02)	.01 (.01)	-.01 (-.01)	.02 (.02)	-.03 (-.03)	.04 (.04)	.00 (.00)	-.07 (-.07)	.10 (.10)	-.01 (-.01)
Dominance	.10 (.08)	.00 (.00)	-.03 (-.03)	-.01 (-.01)	.04 (.05)	.01 (.02)	.01 (.00)	.02 (.01)	-.10 (-.09)	-.02 (-.02)
Even Tempered	.09 (.08)	.09 (.09)	.08 (.07)	.08 (.09)	-.08 (-.07)	.10 **(.10)**	.06 (.06)	.04 (.03)	.01 (.02)	-.01 (-.01)
Generosity	-.07 (-.06)	.07 (.07)	.01 (.01)	**.11 (.11)**	.03 (.02)	.06 (.06)	.03 (.04)	-.02 (-.01)	.01 (.00)	.01 (.01)
Intellectual Efficiency	**.18 (.12)**	-.01 (-.01)	.03 (-.01)	.01 (.02)	-.02 (.01)	.01 (.01)	.05 (.03)	.07 (.02)	**-.17 (-.13)**	-.01 (.00)
Non-delinquency	.02 (.02)	.04 (.04)	.02 (.02)	.01 (.01)	.00 (-.01)	.01 (.01)	.03 (.03)	.02 (.02)	.05 (.05)	.00 (.00)
Optimism	**.12 (.11)**	.02 (.02)	.00 (.00)	.05 (.05)	-.04 (-.04)	.04 (.04)	.00 (.00)	.06 (.06)	-.01 (-.01)	**-.05 (-.05)**
Order	.02 (.05)	.02 (.02)	-.01 (.00)	.03 (.03)	-.03 (-.04)	.06 (.06)	.04 (.05)	-.09 (-.07)	-.02 (-.04)	-.02 (-.02)
Physical Conditioning	**.13 (.12)**	.00 (.00)	-.01 (-.01)	.02 (.02)	-.05 (-.04)	-.02 (-.02)	.05 (.05)	.07 (.06)	-.06 (-.05)	**-.04** (-.04)
Self-Control[a]	-.03 (-.03)	**.14 (.14)**	.03 (.03)	.05 (.05)	-.01 (-.01)	.04 (.04)	.07 (.07)	.00 (.00)	**.14 (.14)**	-.01 (-.01)
Sociability	.05 (.06)	.04 (.04)	-.03 (-.02)	.03 (.03)	.01 (.01)	.07 (.07)	.06 (.06)	.10 **(.11)**	.04 (.03)	-.03 (-.03)
Tolerance	.03 (.04)	**.11 (.11)**	.09 (.10)	.08 (.08)	-.08 (-.08)	.10 (.10)	.10 **(.10)**	.07 (.07)	.02 (.02)	.00 (-.01)
TAPAS Composites										
Can-do Composite	**.19 (.16)**	.08 (.08)	.09 (.07)	.08 (.09)	-.10 (-.09)	.08 (.09)	.09 (.08)	**.11** (.09)	-.06 (-.04)	-.02 (-.02)
Will-do Composite	**.14 (.14)**	.08 (.08)	**.10** (.10)	**.10 (.10)**	**-.11 (-.11)**	.09 (.09)	.09 (.09)	.09 (.09)	-.05 (-.05)	-.02 (-.01)

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in parentheses are semi-partial correlations between the TAPAS scales and the criterion of interest, controlling for AFQT. Estimates in bold were statistically significant, $p < .05$ (two-tailed).

[a] Adjustment and Self Control were included in the TAPAS 15-dimension versions (i.e., static and CAT) only. Sample sizes for these scales are smaller, ranging from 272 – 2,443.

[b] Attrition results include Regular Army Soldiers only.

Table D.5. Correlations between TAPAS Can-do Composite Scores and Performance- and Retention-related Criteria

Criterion	TAPAS Version						AFQT Category (All TAPAS Versions)							
	CAT13		15D		CAT15		I-II		IIIA		IIIB		I-IV	
	r	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*	*r*	*n*
Can-do Performance														
WTBD JKT	.11	87	.01	195	-.03	60	.00	162	.18	78	-.21	80	.03	342
MOS-Specific JKT	-.12	71	.00	161	-.04	42	-.06	134	-.23	60	-.15	64	-.03	274
MOS Proficiency (PRS)	-.06	43	.07	95	--	23	-.03	80	.17	38	.05	35	.03	161
MOS-Specific PRS	-.09	50	-.17	91	--	22	.00	80	-.24	43	-.05	31	-.11	163
IMT Exam Grade	.14	116	.03	350	.00	194	.01	310	-.11	150	.14	176	.04	660
# of Restarts (ALQ)	-.05	380	.05	626	-.09	44	.01	454	.02	267	-.01	287	.01	1,050
Graduated IMT without Restart	.04	380	-.02	626	.25	44	-.04	454	.05	267	.02	287	.00	1,050
Training Achievement (ALQ)	-.18	89	.13	208	.15	64	.10	168	.14	86	.07	85	.07	361
Training Failure (ALQ)	-.01	89	-.07	208	-.17	64	.00	168	-.16	86	-.04	85	-.08	361
Common/Warrior Tasks KS (PRS)	.07	47	.16	98	--	25	.14	84	.14	42	.15	35	.12	170
Will-do Performance														
Exhibiting Effort (PRS)	.09	47	.09	101	-.08	26	-.04	88	.13	42	.25	35	.07	174
Support for Peers (PRS)	.04	49	.07	100	-.04	26	.05	88	.03	43	.07	35	.05	175
Peer Leadership (PRS)	.04	47	.14	93	--	25	-.01	81	.17	41	.10	35	.07	165
Exhibiting Fitness and Bearing (PRS)	.05	49	.17	100	.11	26	.14	89	.16	43	.02	35	.14	175
Personal Discipline (PRS)	.19	49	.11	101	-.11	26	.07	89	.08	43	.18	35	.10	176
Last APFT Score (ALQ)	-.09	88	.01	205	**.25**	64	-.03	166	**.22**	85	-.04	85	.02	357
Disciplinary Action (ALQ)	-.18	47	**-.21**	106	--	23	-.09	81	-.10	44	-.19	40	-.14	176
Commitment and Adjustment (PRS)	.10	49	.09	101	.01	26	.11	89	-.02	43	.16	35	.08	176
Retention														
Adjustment to Army Life (ALQ)	**.21**	89	**.18**	208	.17	64	.08	168	**.25**	86	.12	85	.19	361
Affective Commitment (ALQ)	**.21**	89	.05	208	.00	64	-.03	168	.19	86	.18	85	.08	361
Normative Commitment (ALQ)	**.27**	89	.08	208	-.08	64	.07	168	.16	86	.06	85	.09	361
Career Intentions (ALQ)	.14	89	.05	208	.10	64	.11	168	.05	86	.08	85	.08	361
Attrition Cognitions (ALQ)	**-.22**	89	-.12	208	.12	64	-.09	168	-.14	86	-.02	85	-.10	361
Reenlistment Intentions (ALQ)	.17	89	.05	208	.07	64	.07	168	.07	86	.15	85	.08	361
Army Fit (ALQ)	.19	89	.09	208	-.03	64	-.05	168	**.26**	86	.10	85	.09	361
MOS Fit (ALQ)	.15	89	**.18**	208	-.15	64	.05	168	**.22**	86	-.02	85	.11	361
Army Civilian Comparison (ALQ)	-.10	89	-.05	208	-.09	64	-.05	168	-.04	86	-.15	85	-.06	361
3-Month Attrition	-.04	367	-.01	1,680	-.04	763	.00	1,334	**-.09**	648	-.02	733	-.02	2,810

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in bold were statistically significant, $p < .05$ (two-tailed). 3-month attrition results include Regular Army Soldiers only. Correlation analyses with 25 or fewer cases were suppressed, as represented by (--).

Table D.6. Correlations between TAPAS Will-do Composite Scores and Performance- and Retention-related Criteria

Criterion	TAPAS Version						AFQT Category (All TAPAS Versions)							
	CAT13		15D		CAT15		I-II		IIIA		IIIB		I-IV	
	r	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*
Can-do Performance														
WTBD JKT	.06	87	-.04	195	-.17	60	-.01	162	.14	78	-.18	80	-.04	342
MOS-Specific JKT	-.10	71	-.01	161	-.12	42	-.13	134	-.05	60	.02	64	-.04	274
MOS Proficiency (PRS)	.04	43	.02	95	.10	23	-.11	80	.31	38	.20	35	.03	161
MOS-Specific PRS	.15	50	-.04	91	.03	22	.04	80	-.08	43	.07	31	.01	163
IMT Exam Grade	.01	116	.01	350	-.09	194	-.04	310	-.07	150	.04	176	-.03	660
# of Restarts (ALQ)	-.08	380	.00	626	-.17	44	-.01	454	-.05	267	-.05	287	-.03	1,050
Graduated IMT without Restart	.06	380	.00	626	.13	44	-.02	454	.07	267	.03	287	.02	1,050
Training Achievement (ALQ)	.03	89	.08	208	.09	64	.04	168	.05	86	.15	85	.07	361
Training Failure (ALQ)	-.05	89	-.12	208	.00	64	-.09	168	-.05	86	-.02	85	-.08	361
Common/Warrior Tasks KS (PRS)	.08	47	.10	98	--	25	.00	84	.19	42	.24	35	.09	170
Will-do Performance														
Exhibiting Effort (PRS)	.20	47	.12	101	.00	26	-.05	88	.25	42	**.42**	35	.12	174
Support for Peers (PRS)	.11	49	.07	100	.08	26	.06	88	.08	43	.10	35	.08	175
Peer Leadership (PRS)	.12	47	.10	93	--	25	-.07	81	.24	41	.10	35	.07	165
Exhibiting Fitness and Bearing (PRS)	.16	49	**.20**	100	.20	26	.10	89	.28	43	.22	35	.20	175
Personal Discipline (PRS)	.13	49	.09	101	-.02	26	.01	89	.06	43	.23	35	.09	176
Last APFT Score (ALQ)	**.23**	88	.11	205	.02	64	.11	166	**.26**	85	.09	85	.12	357
Disciplinary Action (ALQ)	-.24	47	**-.30**	106	--	23	-.14	81	-.12	44	-.22	40	-.19	176
Commitment and Adjustment (PRS)	.16	49	.15	101	.09	26	.12	89	.09	43	.28	35	.14	176
Retention														
Adjustment to Army Life (ALQ)	.15	89	**.16**	208	.06	64	.11	168	.07	86	.21	85	.14	361
Affective Commitment (ALQ)	.17	89	.07	208	-.02	64	.02	168	.05	86	.21	85	.08	361
Normative Commitment (ALQ)	**.27**	89	.10	208	-.09	64	**.16**	168	.00	86	.06	85	.10	361
Career Intentions (ALQ)	.12	89	.08	208	.18	64	.14	168	.04	86	.09	85	.10	361
Attrition Cognitions (ALQ)	-.19	89	**-.17**	208	.16	64	**-.15**	168	-.02	86	-.08	85	-.11	361
Reenlistment Intentions (ALQ)	.15	89	.04	208	.17	64	.10	168	-.03	86	.10	85	.09	361
Army Fit (ALQ)	.15	89	.09	208	-.02	64	.01	168	.05	86	.21	85	.09	361
MOS Fit (ALQ)	.09	89	**.14**	208	-.07	64	.03	168	.11	86	.14	85	.09	361
Army Civilian Comparison (ALQ)	.01	89	-.07	208	-.08	64	-.07	168	-.07	86	-.09	85	-.05	361
3-Month Attrition	.03	367	-.02	1,680	-.03	763	.02	1,334	-.07	648	-.03	733	-.02	2,810

Note. AFQT = Armed Forces Qualification Test, TAPAS = Tailored Adaptive Personality Assessment System. ALQ = Army Life Questionnaire. JKT = Job Knowledge Test. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Estimates in parentheses are partial correlations between the TAPAS scales and the criteria of interest controlling for AFQT Category. Estimates in bold were statistically significant, $p < .05$ (two-tailed). 3-month attrition results include Regular Army Soldiers only. Correlation analyses with 25 or fewer cases were suppressed, as represented by (--).

Table D.7. Mean TAPAS Scores for the Target and Expanded Sample of MOS

TAPAS Scale	11B	19K	25U	31B	42A	68W	88M	91B	21B	35F	92G
Achievement	.07	-.03	-.06	.04	-.04	.11	.01	-.08	-.02	.10	-.16
Adjustment	.17	.26	.04	.01	-.17	.08	.02	.05	.12	.04	-.09
Attention Seeking	.10	.06	-.06	.07	-.03	.16	-.01	-.12	.08	.07	-.18
Cooperation	-.01	.10	.19	-.01	.07	.09	.03	-.05	.06	.05	.14
Dominance	.08	-.01	-.10	.13	-.08	.06	-.07	-.18	-.09	.14	-.21
Even Tempered	.02	.14	.04	-.06	.01	.14	.00	.01	.11	.19	.06
Generosity	-.12	-.11	-.01	-.02	.19	.17	-.01	-.04	-.22	-.12	.22
Intellectual Efficiency	.03	-.18	.01	-.07	-.15	.28	-.11	-.16	-.03	.41	-.17
Non-Delinquency	-.01	.05	.11	.11	.05	.06	-.01	-.06	-.07	.21	.14
Optimism	.14	.07	.00	.13	.16	.09	.08	.11	.14	.09	-.07
Order	-.14	-.17	-.09	-.13	.04	-.14	.00	.01	-.20	-.11	.08
Physical Conditioning	.28	-.06	.02	.20	-.10	.04	-.01	.01	.08	-.07	-.17
Self-Control	-.04	.06	-.06	-.04	-.02	-.11	-.02	-.11	-.10	.09	.04
Sociability	.06	-.01	-.02	.07	.03	.06	.04	-.03	.09	-.16	-.01
Tolerance	-.16	-.11	-.10	-.14	.12	.10	-.06	-.17	-.23	.02	.13
TAPAS Composites											
Can-Do Composite	.09	.02	.04	.06	.00	.25	-.01	-.07	.05	.36	-.07
Will-Do Composite	.11	.01	.06	.09	-.03	.07	.00	.00	.01	.14	.02

Note. Results are limited to the Accession Sample (non-prior service, Education Tier 1, AFQT Category IV and above, signed contract). Standardized TAPAS scores were used in this analysis (see Chapter 3). Sample sizes by MOS are: 11B = 2,107, 19K =158, 25U=290, 31B=907, 42A = 410, 68W=1139, 88M = 1,149, 91B = 775, 21B = 572, 35F =338, 92G = 487.

D-7